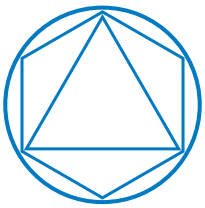
Pair-copula constructions for non-Gaussian Bayesian networks

Alexander Xaver Bauer



Fakultät für Mathematik
Technische Universität München
85748 Garching

2013

TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Mathematische Statistik

Pair-copula constructions for non-Gaussian Bayesian networks

Alexander Xaver Bauer

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Matthias Scherer
Prüfer der Dissertation:	1.	Univ.-Prof. Claudia Czado, Ph.D.
	2.	Prof. Roger M. Cooke, Ph.D. Technische Universiteit Delft, Niederlande
	3.	Prof. Harry Joe, Ph.D. University of British Columbia, Kanada

Die Dissertation wurde am 12.12.2012 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 13.02.2013 angenommen.

To Monika

Zusammenfassung

Pair-Copula-Bayes-Netze (PCBNs) stellen eine neuartige Klasse multivariater statistischer Modelle dar, welche sowohl nichtgaußsche Verteilungen als auch die Vorgabe einer durch einen azyklische Digraphen (ADG) repräsentierten Liste bedingter Unabhängigkeiten erlauben. Die Likelihoodfunktionen dieser Modelle besitzen eine spezifische Faktorisierung, welche auf Pair-Copula-Konstruktionen basiert und folglich nur univariate Verteilungen und bivariate (bedingte) Copulas umfasst.

Wir entwickeln die theoretischen Grundlagen zur Konstruktion von PCBNs auf beliebigen ADGs und liefern generische Algorithmen zur Erzeugung von Zufallsstichproben und zur Likelihood-Inferenz. Darüber hinaus schlagen wir ein Verfahren zur Ordnung der Elternknoten im zugrunde liegenden ADG vor. Wir demonstrieren die Maximum-Likelihood-Schätzung der Parameter dieser Modelle und vergleichen PCBNs mit verschiedenen Modellen aus der Literatur. Eine Simulationsstudie untersucht die Effekte von Modellmissspezifikation und unterstreicht den Bedarf an nichtgaußschen Verteilungen zur Modellierung bedingter Unabhängigkeiten.

Die Schätzung des zugrunde liegenden ADGs erfolgt mit Hilfe einer Variante des bekannten PC-Algorithmus. Diese Variante basiert auf einem neuartigen Test auf bedingte Unabhängigkeit, welcher speziell auf den Modellrahmen zugeschnitten ist. Eine Simulationsstudie zeigt, dass der PC-Algorithmus zur Strukturschätzung in nichtgaußschen PCBNs höchst geeignet ist. Zuletzt untersuchen wir Strategien zur Reduzierung des numerischen Aufwands und verwenden die vorgeschlagenen Verfahren zur Modellierung von Finanzdaten.

Abstract

Pair-copula Bayesian networks (PCBNs) are a novel class of multivariate statistical models which permit non-Gaussian distributions as well as the inclusion of conditional independence assumptions specified by a directed acyclic graph (DAG). These models feature a specific factorisation of the likelihood that is based on pair-copula constructions (PCCs) and hence involves only univariate distributions and bivariate copulas, of which some may be conditional.

We develop the theoretical foundation for constructing PCBNs from arbitrary DAGs and provide generic algorithms for random sampling and likelihood inference. Moreover, routines for selecting orderings of the parents of the vertices in the underlying graphs are proposed. We demonstrate maximum likelihood estimation of the parameters of these models and compare PCBNs to various competing models from the literature. A simulation study investigates the effects of model misspecification and highlights the need for non-Gaussian conditional independence models.

Model selection of the DAG is facilitated using a version of the well-known PC algorithm which is based on a novel test for conditional independence of random variables tailored to the PCC framework. A simulation study shows the PC algorithm's high aptitude for structure estimation in non-Gaussian PCBNs. We finally investigate strategies to reduce computational complexity and apply the proposed methods to modelling financial return data.

Acknowledgements

First of all, I want to thank my supervisors Prof. Claudia Czado and Dr. Thomas Klein for their invaluable advice and encouragement throughout the last years. Prof. Czado generously took over my supervision following Dr. Klein's departure from academia. I am very grateful for the many opportunities they gave me to attend scientific conferences and meet renowned researchers from all over the world.

Moreover, I want to thank my colleagues at the Chair of Mathematical Statistics for their fruitful collaboration, for many stimulating discussions, and for the great time we spent together. I am particularly indebted to Dr. Peter Hepperger for his help with the `C++` implementation of some of the algorithms developed in this thesis.

Also, I gratefully acknowledge the financial support of the TUM Graduate School's Faculty Graduate Center ISAM (International School of Applied Mathematics).

Last, and most importantly, I sincerely thank my wife, Monika, for her infinite patience and loving support.

Contents

Introduction

The analysis of dependence and independence in a given family of random variables is a fundamental concept in multivariate statistics. Although, in a broad sense, independence can be deemed the absence of dependence, inference with a view towards the one or the other may involve substantially different statistical models. In this thesis, we combine the advantages of graphical independence models with the flexibility of copulas to derive a novel multivariate statistical model, which not only reflects Markov properties exhibited by data, but also enables us to capture further distributional features such as heavy-tailedness and non-linear, asymmetric dependence.

Graphical models are multivariate statistical models in which the joint distribution of a family of random variables is restricted by a list of conditional independence assumptions. This list can be summarised in a graph whose vertices represent the variables and whose edges represent interrelations of these variables. We are particularly interested in the graphical models known as Bayesian networks, whose Markov properties can be represented by a directed acyclic graph (DAG). Given a family $\boldsymbol{X} = (X_v)_{v\in\{1,\dots,d\}}$, $d \in \mathbb{N}$, of continuous random variables associated to a DAG $\mathcal{D}$, the distribution of $\boldsymbol{X}$ features a specific factorisation of the probability density function (pdf), which is given by

$$f(\boldsymbol{x}) = \prod_{v=1}^{d} f_{v|\mathrm{pa}(v)}(x_v \,|\, \boldsymbol{x}_{\mathrm{pa}(v)}), \quad \boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d,$$

where $\mathrm{pa}(v)$ are the parents of a vertex v (i.e. the vertices w such that $\mathcal{D}$ contains the edge $w \to v$) and $f_{v|\mathrm{pa}(v)}(\,\cdot \,|\, \boldsymbol{x}_{\mathrm{pa}(v)})$ is the conditional pdf of X_v given $\boldsymbol{X}_{\mathrm{pa}(v)} = \boldsymbol{x}_{\mathrm{pa}(v)}$ (Lauritzen, 1996, Chapter 3). For instance, if $\boldsymbol{X}$ follows a multivariate normal

distribution, this factorisation is reflected in a particular shape of the precision matrix. If $\boldsymbol{X}$ follows a discrete distribution, a similar factorisation holds for the probability mass function. Inference in Bayesian networks is thus facilitated by their visual representation of conditional independence restrictions through missing edges in a graph and by the factorisation of their likelihood based on parent-child relationships. Moreover, the inclusion of Markov properties results in a certain parsimony.

The origins of graphical models can be traced back to the seminal work of Wright (1921) on path analysis. Modern applications of Bayesian networks range from artificial intelligence, decision support systems, and engineering to genetics, geology, medicine, and finance (Pourret et al., 2008). These models play an important role in causal reasoning (Pearl, 2009). Despite the broad scope of applicability, however, graphical modelling of continuous random variables has mainly been limited to the multivariate normal distribution. At the same time, it is well known from the literature on statistical finance that the assumption of joint normality may lead to severe underestimation of certain risks (McNeil et al., 2005), and, in a more general sense, fails to yield suitable models in many applications. Another approach to dealing with continuous Bayesian networks is to discretise the continuous random variables and then to work with the corresponding discrete model. However, depending on the discretisation used, the conditional probability tables involved in this procedure quickly become very complex. We instead use copulas to derive flexible and tractable continuous Bayesian networks that are not hampered by these drawbacks.

Copulas are multivariate cumulative distribution functions (cdfs) on the unit cube with uniform univariate margins. Their popularity originates from a theorem of Sklar (1959), which states that every multivariate cdf can be decomposed into its univariate margins and a copula. Given a family $\boldsymbol{X} = (X_v)_{v\in\{1,\dots,d\}}$ of continuous random variables with strictly increasing univariate marginals $F_1, \dots, F_d$, this relationship can also be stated in terms of the pdf as

$$f(\boldsymbol{x}) = c\big(F_1(x_1), \dots, F_d(x_d)\big) \prod_{v=1}^{d} f_v(x_v), \quad \boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d,$$

where the copula pdf c is uniquely determined. From a statistical perspective, copulas allow us to separate the tasks of modelling univariate margins and multivariate dependences. Their high usability has lead to widespread application, in particular in quantitative finance (Cherubini et al., 2004). While there is a plethora of literature on bivariate copulas (also called pair copulas) (Joe, 1997; Nelsen, 2006), however, many of these bivariate copulas have no straightforward multivariate extension.

Bedford and Cooke (2001, 2002) hence developed pair-copula constructions (PCCs), in which a multivariate copula is decomposed into bivariate, potentially conditional, copulas. Also, they introduced so-called regular vines (R-vines) as a graphical representation of a class of hierarchical PCC models known as vine copula models. Aas et al. (2009) and Dißmann et al. (2013) later recognised these models' aptitude for likelihood inference since pdfs of vine copulas are easily obtainable in explicit analytical form. By drawing on the rich literature on pair-copula families, vine copula models provide high distributional flexibility and can thus be universally applied in multivariate dependence analysis. Applications to financial data have shown that vine copula models outperform other multivariate copula models in predicting log-returns of equity portfolios, see Aas and Berg (2009), Chollete et al. (2009), and Fischer et al. (2009). While one may be reminded of graphical models when looking at R-vines, these graphical representations have no interpretation in terms of Markov properties. In fact, not all Markov properties specified by a DAG can be represented in a vine copula model. Also, the flexibility of vine copula models comes at the price of quadratic growth in the number of pair copulas to be specified as the number of variables increases.

In an attempt to merge the advantages of both worlds, Kurowicka and Cooke (2005) applied the pair-copula concept to Bayesian networks and thereby obtained a novel class of PCC models which, by construction, satisfy conditional independence assumptions specified by a DAG. However, their analyses were restricted to pair-copula families with the property that zero rank correlation implies independence. We cast their idea into a more general theoretical framework and show that this restriction of the set of applicable pair copulas is unnecessary. Given a family $\boldsymbol{X} = (X_v)_{v \in \{1,\dots,d\}}$

of continuous random variables with strictly increasing univariate marginals and Markov structure specified by a DAG $\mathcal{D}$, iterative application of Sklar's theorem to the pdf factorisation with respect to $\mathcal{D}$ yields the decomposition

$$f(\boldsymbol{x}) = \prod_{v=1}^{d} f_v(x_v) \prod_{w \in \mathrm{pa}(v)} c_{v,w|\mathrm{pa}(v;w)}\big(F_{v|\mathrm{pa}(v;w)}(x_v|\boldsymbol{x}_{\mathrm{pa}(v;w)}), F_{w|\mathrm{pa}(v;w)}(x_w|\boldsymbol{x}_{\mathrm{pa}(v;w)})\big|\boldsymbol{x}_{\mathrm{pa}(v;w)}\big),$$

where $\mathrm{pa}(v;w)$ are the parents of v that are smaller than w by a given increasing ordering $<_v$ of $\mathrm{pa}(v)$. The parent orderings $<_v$ need to be specified in addition to the Markov structure. This way, we obtain a pdf decomposition in which each edge of $\mathcal{D}$ corresponds to exactly one (conditional) pair copula. From a statistical viewpoint, we obtain a novel statistical model that is both a Bayesian network and a copula model, and in which all copula parameters can be chosen independently of each other. We refer to these models, which include Gaussian Bayesian networks as a special case, as *pair-copula Bayesian networks* (PCBNs).

The price to be paid for the inclusion of conditional independence assumptions is reflected in the conditional cdfs, whose evaluation may require high-dimensional integration and is thus potentially more involved than in vine copulas. So far, an explicit representation of these conditional cdfs was only derived in examples. We provide and proof a general algorithm for evaluating the likelihood of an arbitrary PCBN. Further research on PCBNs includes Hanea et al. (2006) and Hanea and Kurowicka (2008). While these authors concentrate on non-parametric statistical inference and elicited expert knowledge, we focus attention to parametric likelihood inference and data-driven structure estimation. Elidan (2010) uses another copula decomposition of distributions associated with a DAG that is based on generally higher-variate copulas and hence lacks the flexibility of the pair-copula approach. Overall, we observe that copulas are slowly starting to make their way into graphical modelling.

A concept crucial to PCCs are conditional copulas. A customary assumption in PCC modelling is that families and parameters of conditional pair copulas remain constant for all possible values of the conditioning variables, see Hobæk Haff et al. (2010), Acar et al. (2012), and Stöber et al. (2012) for a critical assessment. Outside the PCC

context, conditional copulas have been used in finance applications, frequently with the aim of modelling time-varying dependence (Patton, 2006). Here, time variation in the conditional copulas is captured by the inclusion of time-varying parameters.

Model selection in PCBNs involves specification of the underlying DAG. When expert knowledge is unavailable, data-driven structure estimation algorithms are frequently used. Two approaches are predominantly found in the literature: the constraint-based and the score-and-search-based approach (Koller and Friedman, 2009, Chapter 18). In the former, the DAG is inferred from a series of conditional independence tests, while in the latter, the DAG is found by optimising a given scoring function. Unfortunately, available implementations of aforementioned algorithms are mainly confined to discrete or Gaussian models. We concentrate on the constraint-based PC algorithm (Spirtes and Glymour, 1991) and introduce a novel test for conditional independence of continuous random variables tailored to the PCC framework. This test uses the Rosenblatt transform (Rosenblatt, 1952) to convert the problem of testing the conditional independence $X_u \perp\!\!\!\perp X_v \mid \boldsymbol{X}_W$, $u, v \notin W \subseteq \{1, \dots, d\}$, into the problem of testing the independence $F_{u|W}(X_u \mid \boldsymbol{X}_W) \perp\!\!\!\perp F_{v|W}(X_v \mid \boldsymbol{X}_W)$. The conditional cdfs required in this transformation are estimated using vine copula models.

Outline of the thesis

In Chapter 1, we review Bayesian networks. We begin in Section 1.1 by collecting the graph-theoretical terminology required throughout this thesis. We then give the definition of a $\mathcal{D}$-Markovian probability measure in Section 1.2 and state some important properties. In Section 1.3, we introduce graphical models and focus attention to Gaussian Bayesian networks.

Vine copula models are introduced in Chapter 2. We first recall the definition of a copula in Section 2.1 and provide the cdfs and pdfs of the pair-copula families used in this thesis. Next, we introduce PCCs in Section 2.2, before considering likelihood inference and model selection in vine copula models in Section 2.3.

In Chapter 3, we develop the theoretical foundation for constructing PCBNs and investigate statistical inference in these models. Above given pair-copula decomposition of distributions associated to a DAG is proved in Section 3.1. Next, we formally derive pair-copula factorisations for the corresponding conditional cdfs in Section 3.2 and summarise the results in a ready-to-use algorithm. We proceed by introducing PCBNs in Section 3.3, where we consider random sampling as well as likelihood inference. The developed routines for likelihood inference are eventually tested in a simulation study in Section 3.4, which also investigates the effects of model misspecification. It is demonstrated that non-Gaussian PCBNs outperform Gaussian PCBNs and D-vine copula models in all of the investigated settings. Sections 3.1 and 3.4 are based on Bauer et al. (2012), while Sections 3.2 and 3.3 are taken from Bauer and Czado (2012).

Model selection in PCBNs is studied in Chapter 4. We begin by considering structure estimation in PCBNs in Section 4.1. To this end, we review the PC algorithm and introduce the novel test for conditional independence of continuous random variables mentioned above. We then examine the PC algorithm's aptitude for structure estimation in non-Gaussian PCBNs in a simulation study in Section 4.2. Our novel test will prove to outperform a standard test for zero partial correlation frequently used in the Gaussian setting. Last, we propose a strategy for ordering the parents of a vertex in the DAG in Section 4.3. This strategy is based on the idea of modelling strongest dependences in the pair copulas with fewest conditioning variables. Chapter 4 is based on Bauer and Czado (2012).

In Chapter 5, the hitherto developed routines are applied to modelling financial return data. More precisely, we model daily log-returns of two portfolios comprising German and US stock and bond indices (Section 5.1) and international stock indices (Section 5.2). In the first data set, we use economic considerations to derive the underlying DAG, while in the second data set, we use the PC algorithm. The chapter is concluded by an account on reducing computational complexity. Section 5.1 is taken from Bauer et al. (2012), while Section 5.2 is based on Bauer and Czado (2012).

The thesis finally contains two appendices. In Appendix A, we provide closed form

1 Bayesian networks

Bayesian networks are multivariate statistical models in which the joint distribution of a family of random variables is restricted by a list of conditional independence assumptions. This list can be summarised in a directed acyclic graph $\mathcal{D}$ whose vertices represent the variables and whose edges represent interrelations between these variables. We begin this chapter by fixing some graph-theoretical terminology, which is not used consistently throughout the literature. We then give the definition of a $\mathcal{D}$-Markovian probability measure and state some important properties. Finally, we introduce graphical models and focus attention to Gaussian Bayesian networks. A comprehensive introduction to Bayesian networks is found in Lauritzen (1996) and Cowell et al. (2003), see also Pourret et al. (2008) for examples of applications.

1.1 Graph-theoretical terminology

Let $V \neq \emptyset$ be a finite set and let $E \subseteq \mathcal{E} \coloneqq \big\{(v,w) \in V \times V \,\big|\, v \neq w\big\}$. Then $\mathcal{G} = (V,E)$ denotes a *graph* with *vertex set* V and *edge set* E. We say that $\mathcal{G}$ contains the *undirected edge* $v - w$ if $(v,w) \in E$ and $(w,v) \in E$. Similarly, we say that V contains the *directed edge* $v \to w$ if $(v,w) \in E$ but $(w,v) \notin E$. A graph containing only undirected edges is called an *undirected graph* (UG). If $\mathcal{G}$ is a UG and $E \equiv \mathcal{E}$, we call $\mathcal{G}$ the *complete UG* on V. Moreover, a graph containing only directed edges is called a *directed graph*. By replacing all directed edges of $\mathcal{G}$ with undirected edges, we obtain the *skeleton* $\mathcal{G}^s$ of $\mathcal{G}$. We write $v \multimap w$ whenever $(v,w) \in E$, that is $\mathcal{G}$ contains either the directed edge $v \to w$ or the undirected edge $v - w$. A sequence of distinct vertices $v_1, \ldots, v_k \in V$, $k \geq 2$, is called a *path* from v_1 to v_k if $\mathcal{G}$ contains $v_i \multimap v_{i+1}$ for all $i \in \{1, \ldots, k-1\}$. A path from v_1 to v_k is called *directed*

if at least one of the connecting edges is directed. We call a path from v_1 to v_k a *cycle* if $v_1 = v_k$. In particular, we call a directed path from v_1 to v_k a *directed cycle* if $v_1 = v_k$. A graph without directed cycles is known as a *chain graph* (CG). A CG containing only directed edges is called a *directed acyclic graph* (DAG). We define the *adjacency set* of a vertex $v \in V$ as $\mathrm{ad}(v) \coloneqq \big\{w \in V \,\big|\, (v,w) \in E \text{ or } (w,v) \in E\big\}$. If $w \notin \mathrm{ad}(v)$, we say that v and w are *non-adjacent*. A triple of vertices (u,v,w) is called a *v-structure* if $\mathcal{G}$ contains $u \to v \leftarrow w$ and if u and w are non-adjacent.

Now let $\mathcal{D} = (V,E)$ be a DAG. The *moral graph* $\mathcal{D}^m$ of $\mathcal{D}$ is defined as the skeleton of the graph obtained from $\mathcal{D}$ by introducing an undirected edge $u - w$ whenever $\mathcal{D}$ contains a v-structure (u,v,w) for $u,v,w \in V$. Since all edges of $\mathcal{D}$ are directed, we can speak of paths instead of directed paths. For $v \in V$, we set

$$\begin{aligned}
\mathrm{pa}(v) &\coloneqq \big\{w \in V \,\big|\, \mathcal{D} \text{ contains } w \to v\big\} && (\textit{parents} \text{ of } v),\\
\mathrm{an}(v) &\coloneqq \big\{w \in V \,\big|\, \mathcal{D} \text{ contains a path from } w \text{ to } v\big\} && (\textit{ancestors} \text{ of } v),\\
\mathrm{de}(v) &\coloneqq \big\{w \in V \,\big|\, \mathcal{D} \text{ contains a path from } v \text{ to } w\big\} && (\textit{descendants} \text{ of } v), \text{ and}\\
\mathrm{nd}(v) &\coloneqq V \setminus \big(\{v\} \cup \mathrm{de}(v)\big) && (\textit{non-descendants} \text{ of } v).
\end{aligned}$$

A set $I \subseteq V$ is called *ancestral* if $\mathrm{pa}(v) \subseteq I$ for all $v \in I$. The smallest ancestral set containing I is denoted by $\mathrm{An}(I)$. As is readily verified, $\mathrm{An}(I) = I \cup \bigcup_{v \in I} \mathrm{an}(v)$. The graph $\mathcal{D}_I = \big(I, E \cap (I \times I)\big)$ is called the *subgraph* of $\mathcal{D}$ induced by I. A bijection $v_\bullet \colon \big\{1,\dots,|V|\big\} \to V$, $i \mapsto v_i$, satisfying $i < j$ whenever $\mathcal{D}$ contains $v_i \to v_j$ for some $i,j \in \big\{1,\dots,|V|\big\}$ is called a *well-ordering* of $\mathcal{D}$. Note that in a well-ordered DAG the set $\{v_1,\dots,v_k\}$ is ancestral for all $k \in \big\{1,\dots,|V|\big\}$.

Finally, let $\mathcal{G}$ be a UG and let $I,J,K \subseteq V$ be pairwise disjoint. A path from I to J is a path from a vertex $v \in I$ to a vertex $w \in J$. We say that K *separates* I from J in $\mathcal{G}$, and write $I \perp J \mid K\,[\mathcal{G}]$, if every path from I to J contains a vertex in K. In particular, we write $I \perp J \mid \emptyset\,[\mathcal{G}]$, or shortly $I \perp J\,[\mathcal{G}]$, if there exists no path between I and J. We call $\mathcal{G}$ *connected* if for every distinct $v,w \in V$ there is a path from v to w. A connected UG without cycles is a *tree*. If there is a vertex $w \in V$ such that $\mathrm{ad}(w) = V \setminus \{w\}$ and $\mathrm{ad}(v) = \{w\}$ for all $v \in V \setminus \{w\}$, that is all vertices

are solely adjacent to w, then $\mathcal{G}$ is called a *star* and w is called its *root vertex*.

Example 1.1. Figure 1.1 shows a well-ordered DAG $\mathcal{D}$ on four vertices $1, 2, 3, 4$ together with the corresponding moral graph $\mathcal{D}^m$. For every vertex v in $\mathcal{D}$, the sets $\mathrm{ad}(v)$, $\mathrm{an}(v)$, $\mathrm{de}(v)$, $\mathrm{nd}(v)$, and $\mathrm{pa}(v)$, respectively, are given in Table 1.1. Observe that $\mathcal{D}$ contains the v-structure $2 \to 4 \leftarrow 3$, which is why $\mathcal{D}^m$ contains the edge $2 - 3$. We obtain the graph separation property $1 \perp 4 \mid \{2, 3\}\, [\mathcal{D}^m]$.

1.2 Conditional independence and directed acyclic graphs (DAGs)

In graphical probability modelling, graphs are used to represent conditional independence properties of corresponding families of probability measures. Let $\mathcal{D} = (V, E)$ be a DAG on $d \coloneqq |V|$ vertices and let P be a probability measure on $\mathbb{R}^d$. Moreover, let $\boldsymbol{X}$ be an $\mathbb{R}^d$-valued random variable distributed as P. For $I \subseteq V$, we write $\boldsymbol{X}_I \coloneqq (X_v)_{v \in I}$ and denote the corresponding I-margin of P by P_I. If $I = \{v\}$ for some $v \in V$, we write X_v and P_v instead of $X_{\{v\}}$ and $P_{\{v\}}$. Furthermore, we write $\boldsymbol{X}_I \perp\!\!\!\perp \boldsymbol{X}_J \mid \boldsymbol{X}_K$ whenever $\boldsymbol{X}_I$ and $\boldsymbol{X}_J$ are *conditionally independent* given $\boldsymbol{X}_K$ for pairwise disjoint sets $I, J, K \subseteq V$. By convention, $\boldsymbol{X}_I \perp\!\!\!\perp \boldsymbol{X}_J \mid \boldsymbol{X}_\emptyset$ is understood as $\boldsymbol{X}_I \perp\!\!\!\perp \boldsymbol{X}_J$. P is said to possess the *local $\mathcal{D}$-Markov* property if

$$X_v \perp\!\!\!\perp \boldsymbol{X}_{\mathrm{nd}(v) \setminus \mathrm{pa}(v)} \mid \boldsymbol{X}_{\mathrm{pa}(v)} \quad \text{for all } v \in V. \tag{1.1}$$

Correspondingly, P is said to possess the *global $\mathcal{D}$-Markov* property if

$$I \perp J \mid K\, [(\mathcal{D}_{\mathrm{An}(I \cup J \cup K)})^m] \;\Rightarrow\; \boldsymbol{X}_I \perp\!\!\!\perp \boldsymbol{X}_J \mid \boldsymbol{X}_K \quad \text{for all } I, J, K \subseteq V. \tag{1.2}$$

Equations (1.1) and (1.2) relate (conditional) independence properties of P to graph separation properties of $\mathcal{D}$. Since $\mathrm{ad}(v) \cap \big(\mathrm{nd}(v) \setminus \mathrm{pa}(v)\big) = \emptyset$ for every $v \in V$, it can be easily seen that the conditional independence restrictions obtained from Equation (1.1) correspond to missing edges in $\mathcal{D}$. One can show that P has the local $\mathcal{D}$-Markov property if and only if P has the global $\mathcal{D}$-Markov property, see Lauritzen (1996,

p. 51). A probability measure satisfying Equations (1.1) and (1.2) is thus simply called $\mathcal{D}$-*Markovian*. Despite the aforementioned equivalence, the lists of *explicit* conditional independence restrictions obtained from Equations (1.1) and (1.2) may, however, be of different lengths. Note that a $\mathcal{D}$-Markovian probability measure can exhibit further conditional independence properties apart from those represented by $\mathcal{D}$. If, however, P exhibits no conditional independence properties other than those represented by $\mathcal{D}$, then P is called *faithful* to $\mathcal{D}$. Now let P have Lebesgue-density f. One can show that P is $\mathcal{D}$-Markovian if and only if f has a so-called $\mathcal{D}$-*recursive factorisation*, that is

$$f(\boldsymbol{x}) = \prod_{v \in V} f_{v|\mathrm{pa}(v)}\big(x_v \,\big|\, \boldsymbol{x}_{\mathrm{pa}(v)}\big) \quad \text{for all } \boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d, \tag{1.3}$$

where $f_{v|\mathrm{pa}(v)}(\,\cdot \mid \boldsymbol{x}_{\mathrm{pa}(v)})$ denotes the conditional probability density function (pdf) of X_v given $\boldsymbol{X}_{\mathrm{pa}(v)} = \boldsymbol{x}_{\mathrm{pa}(v)}$, see again Lauritzen (1996, p. 51). Note that there may be more than one DAG representing the same set of conditional independence restrictions. We call the set of DAGs representing the same conditional independence restrictions as $\mathcal{D}$ the *Markov-equivalence class* of $\mathcal{D}$, and denote it by $[\mathcal{D}]$. Two DAGs $\mathcal{D}_1 = (V, E_1)$ and $\mathcal{D}_2 = (V, E_2)$ are called *Markov equivalent* if $[\mathcal{D}_1] = [\mathcal{D}_2]$. By Verma and Pearl (1991), $\mathcal{D}_1$ and $\mathcal{D}_2$ are Markov equivalent if and only if they have the same skeleton and the same v-structures. The Markov-equivalence class of $\mathcal{D}$ can be represented by a CG, the so-called *essential graph* $\mathcal{D}^e$ associated with $[\mathcal{D}]$, which has the same skeleton as $\mathcal{D}$ and contains a directed edge $v \to w$ if and only if all members of $[\mathcal{D}]$ contain $v \to w$, see Andersson et al. (1997). A DAG in $[\mathcal{D}]$ can be obtained from $\mathcal{D}^e$ by directing all undirected edges of $\mathcal{D}^e$ such that no new v-structures and no directed cycles are introduced.

Example 1.2. Figure 1.2 shows again the DAG $\mathcal{D}$ from Figure 1.1 together with the essential graph associated with the corresponding Markov-equivalence class. Let P be an absolutely continuous $\mathcal{D}$-Markovian probability measure on $\mathbb{R}^4$. Straightforward evaluation of Condition (1.1) yields the restrictions $X_1 \perp\!\!\!\perp X_\emptyset \mid X_\emptyset$ (for $v = 1$), $X_2 \perp\!\!\!\perp X_3 \mid X_1$ (both for $v = 2$ and $v = 3$), and $X_1 \perp\!\!\!\perp X_4 \mid \boldsymbol{X}_{23}$ (for $v = 4$), of which the first is vacuous. There are no other implicit conditional independence properties

in this example. As for the pdf of P, Equation (1.3) yields the representation

$$f(\boldsymbol{x}) = f_1(x_1)\, f_{2|1}(x_2 \,|\, x_1)\, f_{3|1}(x_3 \,|\, x_1)\, f_{4|23}(x_4 \,|\, \boldsymbol{x}_{23}) \quad \text{for all } \boldsymbol{x} = (x_1, \ldots, x_4) \in \mathbb{R}^4.$$

We shall mention that verifying whether a given empirical distribution on $\mathbb{R}^d$ can be assumed to originate from a $\mathcal{D}$-Markovian probability measure is hard. One approach is to apply structure estimation algorithms to the given data, cf. Chapter 4. As an alternative approach, expert knowledge is frequently exploited to define the graph $\mathcal{D}$ specifying the Markov structure (Kurowicka and Cooke, 2006, Chapter 5).

1.3 Gaussian Bayesian networks

A *Bayesian network* or *(directed) graphical model* based on a DAG $\mathcal{D} = (V, E)$ is a family of $\mathcal{D}$-Markovian probability measures. For instance, the family of all regular $\mathcal{D}$-Markovian normal distributions on $\mathbb{R}^d$, $d := |V|$, is called the *Gaussian Bayesian network* based on $\mathcal{D}$. Applications of Bayesian networks range from artificial intelligence, decision support systems, and engineering to genetics, geology, medicine, and finance. For lack of tractable continuous probability measures, statistical modelling with Bayesian networks has mostly been limited to multivariate discrete or normal distributions. Kurowicka and Cooke (2005) used copulas to derive a rich and tractable class of continuous Bayesian networks, which we investigate in Chapter 3.

Let us briefly review Gaussian Bayesian networks. By $\{\mathrm{N}_{\mathcal{D},\boldsymbol{\Sigma}} \,|\, \boldsymbol{\Sigma} \in \Theta\}$ we denote the family of $\mathcal{D}$-Markovian normal distributions on $\mathbb{R}^d$ with zero mean and positive definite correlation matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{x} = (\boldsymbol{x}^1, \ldots, \boldsymbol{x}^n)$, $n \in \mathbb{N}$, be a realisation of a sample of i.i.d. observations $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^n$ from a corresponding random variable $\boldsymbol{X}$. As is well known, the log-likelihood function takes the form

$$l(\boldsymbol{\Sigma}; \boldsymbol{x}) = -\frac{n\,d}{2} \log(2\pi) - \frac{n}{2} \log|\boldsymbol{\Sigma}| - \frac{1}{2} \operatorname{tr}\big(\boldsymbol{\Sigma}^{-1}\, \boldsymbol{x}^{\mathrm{T}}\, \boldsymbol{x}\big),$$

where $|\boldsymbol{A}|$ denotes the determinant and $\operatorname{tr}(\boldsymbol{A})$ the trace of a matrix $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ (Lauritzen, 1996, Section 5.1.2). A more explicit representation of the precision

Figure 1.1: A well-ordered (vertex labels) DAG $\mathcal{D}$ (left) with moral graph $\mathcal{D}^m$ (right).

v	$\mathrm{ad}(v)$	$\mathrm{an}(v)$	$\mathrm{de}(v)$	$\mathrm{nd}(v)$	$\mathrm{pa}(v)$
1	$\{2,3\}$	$\emptyset$	$\{2,3,4\}$	$\emptyset$	$\emptyset$
2	$\{1,4\}$	$\{1\}$	$\{4\}$	$\{1,3\}$	$\{1\}$
3	$\{1,4\}$	$\{1\}$	$\{4\}$	$\{1,2\}$	$\{1\}$
4	$\{2,3\}$	$\{1,2,3\}$	$\emptyset$	$\{1,2,3\}$	$\{2,3\}$

Table 1.1: Sets of adjacencies (ad), ancestors (an), descendants (de), non-descendants (nd), and parents (pa) for the vertices v in DAG $\mathcal{D}$ of Figure 1.1.

Figure 1.2: A DAG $\mathcal{D}$ (left) specifying the conditional independence restrictions $X_1 \perp\!\!\!\perp X_4 \mid \boldsymbol{X}_{23}$ and $X_2 \perp\!\!\!\perp X_3 \mid X_1$, and the essential graph $\mathcal{D}^e$ (right) associated with the corresponding Markov-equivalence class $[\mathcal{D}]$.

matrix $\boldsymbol{\Sigma}^{-1}$, and hence of $\boldsymbol{\Sigma}$, can be given using the Markov properties of $\mathrm{N}_{\mathcal{D},\boldsymbol{\Sigma}}$. Assume w.l.o.g. that $\mathcal{D}$ is well-ordered, and write $V = \{1, \ldots, d\}$. Let $\boldsymbol{A} \in \mathbb{R}^{d \times d}$ be the lower triangular matrix with elements

$$A_{i,i} \coloneqq 1, \quad \boldsymbol{A}_{i,\mathrm{pa}(i)} \coloneqq -\boldsymbol{\Sigma}_{i,\mathrm{pa}(i)} \left(\boldsymbol{\Sigma}_{\mathrm{pa}(i),\mathrm{pa}(i)}\right)^{-1}, \quad A_{i,j} \coloneqq 0 \quad \text{otherwise,}$$

and let $\boldsymbol{D} \in \mathbb{R}^{d \times d}$ be the diagonal matrix with elements

$$D_{i,i} \coloneqq \Sigma_{i,i} - \boldsymbol{\Sigma}_{i,\mathrm{pa}(i)} \left(\boldsymbol{\Sigma}_{\mathrm{pa}(i),\mathrm{pa}(i)}\right)^{-1} \boldsymbol{\Sigma}_{\mathrm{pa}(i),i}.$$

Then $\boldsymbol{\Sigma}^{-1} = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{D}^{-1} \boldsymbol{A}$, see Cox and Wermuth (1996, Chapter 3).

Moreover, let $\boldsymbol{\Delta} \in \mathbb{R}^{d \times d}$ be the diagonal matrix with elements $\Delta_{i,i} \coloneqq \sqrt{(\Sigma^{-1})_{i,i}}$. The elements of the standardised precision matrix $\boldsymbol{C} \coloneqq \boldsymbol{\Delta}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Delta}^{-1}$ have an interpretation in terms of *partial correlations*. More precisely, let $i \neq j \in V$. Then $C_{i,j} = -\rho_{i,j \cdot V \setminus \{i,j\}}$ (Lauritzen, 1996, Section 5.1.3). In particular, we have

$$C_{i,j} = 0 \quad \Leftrightarrow \quad X_i \perp\!\!\!\perp X_j \mid \boldsymbol{X}_{V \setminus \{i,j\}}.$$

Also, for every $K \subseteq V \setminus \{i, j\}$ and an arbitrary $k \in K$, we have the recursive formula

$$\rho_{i,j \cdot K} = \frac{\rho_{i,j \cdot K_{-k}} - \rho_{i,k \cdot K_{-k}} \, \rho_{j,k \cdot K_{-k}}}{\sqrt{1 - \rho^2_{i,k \cdot K_{-k}}} \sqrt{1 - \rho^2_{j,k \cdot K_{-k}}}}, \tag{1.4}$$

where $K_{-k} \coloneqq K \setminus \{k\}$, see Kurowicka and Cooke (2006, Section 3.3).

2 Vine copula models

Copulas provide a powerful tool in statistical analysis, which allows to conveniently separate the tasks of modelling univariate margins and multivariate dependences. Vine copulas are multivariate copulas which use bivariate copulas as building blocks and which feature a graphical representation called a regular vine. In this chapter, we introduce copulas and give some important examples. Moreover, we review pair-copula constructions and consider statistical inference in vine copula models. A comprehensive introduction to copulas is found in Joe (1997) and Nelsen (2006). Cherubini et al. (2004, 2011) discuss applications of copulas in finance. Vine copula models are treated in Kurowicka and Cooke (2006) and Kurowicka and Joe (2011).

2.1 Copulas

A d-variate *copula*, $d \in \mathbb{N}$, is a cumulative distribution function (cdf) on $[0,1]^d$ such that all univariate marginals are uniform on the interval $[0,1]$. By *Sklar's theorem* (Sklar, 1959), every cdf F on $\mathbb{R}^d$ with marginals $F_1, \dots, F_d$ can be written as

$$F(\boldsymbol{x}) = C\big(F_1(x_1), \dots, F_d(x_d)\big), \quad \boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d,$$

for some suitable copula C. If F is absolutely continuous and $F_1, \dots, F_d$ are strictly increasing, a similar relationship holds for the pdf f of F, namely

$$f(\boldsymbol{x}) = c\big(F_1(x_1), \dots, F_d(x_d)\big) \prod_{i=1}^{d} f_i(x_i), \quad \boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d,$$

where the *copula pdf* c is uniquely determined. Above equations can then be solved for C and c, respectively, using marginal quantile functions. Doing so, we get

$$C(\boldsymbol{u}) = F\big(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\big) \quad \text{and} \quad c(\boldsymbol{u}) = \frac{f\big(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\big)}{\prod_{i=1}^d f_i\big(F_i^{-1}(u_i)\big)} \tag{2.1}$$

for all $\boldsymbol{u} = (u_1, \dots, u_d) \in [0,1]^d$. For instance, consider the case of complete independence. Then Equation (2.1) yields the *independence copula* $C^{\perp\!\!\!\perp}(\boldsymbol{u}) \coloneqq \prod_{i=1}^d u_i$ with pdf $c^{\perp\!\!\!\perp}(\boldsymbol{u}) \equiv 1$.

Two important classes of copulas are the elliptical and the Archimedean copulas. We will use copulas from these two classes in the simulation studies and data applications of Chapters 3 to 5. *Elliptical copulas* arise from applying Equation (2.1) to elliptical distributions, which include the multivariate normal and Student's t distributions. On the other hand, let $\varphi\colon [0,1] \to [0,\infty]$ be a continuous, strictly decreasing, convex function satisfying $\varphi(0) = \infty$ and $\varphi(1) = 0$. Then

$$C^{\varphi}(\boldsymbol{u}) \coloneqq \varphi^{-1}\big(\varphi(u_1) + \cdots + \varphi(u_d)\big) \tag{2.2}$$

is called a *strict Archimedean copula* with *generator* φ. We are mostly interested in *bivariate* elliptical and Archimedean copulas, of which we now give some examples.

Example 2.1 (Gaussian copula)**.** The bivariate *Gaussian copula* with parameter $\rho \in (-1,1)$ is an elliptical copula which arises from applying Equation (2.1) to the bivariate standard normal cdf Φ_ρ with correlation coefficient ρ. We get

$$C(\boldsymbol{u}; \rho) = \Phi_\rho\big(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\big), \quad \boldsymbol{u} = (u_1, u_2) \in [0,1]^2,$$

where Φ denotes the univariate standard normal cdf. Setting $x_1 \coloneqq \Phi^{-1}(u_1)$ and $x_2 \coloneqq \Phi^{-1}(u_2)$, the corresponding pdf takes the form

$$c(\boldsymbol{u}; \rho) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho^2(x_1^2 + x_2^2) - 2\,\rho\, x_1\, x_2}{2\,(1-\rho^2)}\right).$$

Example 2.2 (Student's t copula)**.** The bivariate *Student's t copula* with parameters $\rho \in (-1,1)$ and $\nu \in (1,\infty)$ is another elliptical copula. It arises from applying

Equation (2.1) to the bivariate standard t cdf $\mathrm{t}_{\rho,\nu}$ with correlation coefficient ρ and ν degrees of freedom. We obtain

$$C(\boldsymbol{u};\rho,\nu) = \mathrm{t}_{\rho,\nu}\left(\mathrm{t}_\nu^{-1}(u_1), \mathrm{t}_\nu^{-1}(u_2)\right), \quad \boldsymbol{u} = (u_1, u_2) \in [0,1]^2,$$

where t_ν denotes the univariate Student's t cdf with ν degrees of freedom. Setting $x_1 \coloneqq \mathrm{t}_\nu^{-1}(u_1)$ and $x_2 \coloneqq \mathrm{t}_\nu^{-1}(u_2)$, the corresponding pdf is given by

$$c(\boldsymbol{u};\rho,\nu) = \frac{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\nu+2}{2}\right)\left(1+\frac{x_1^2}{\nu}\right)^{\frac{\nu+1}{2}}\left(1+\frac{x_2^2}{\nu}\right)^{\frac{\nu+1}{2}}}{\sqrt{1-\rho^2}\,\Gamma^2\left(\frac{\nu+1}{2}\right)\left(1+\frac{x_1^2+x_2^2-2\rho x_1 x_2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}},$$

where Γ denotes the gamma function.

Example 2.3 (Clayton copula)**.** The *Clayton copula* with parameter $\theta \in (0,\infty)$ is an Archimedean copula with generator $\varphi_\theta(t) \coloneqq \frac{1}{\theta}\left(t^{-\theta}-1\right)$. Equation (2.2) yields

$$C(\boldsymbol{u};\theta) = \left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-\frac{1}{\theta}}, \quad \boldsymbol{u} = (u_1,u_2) \in [0,1]^2, \quad \text{and}$$

$$c(\boldsymbol{u};\theta) = (1+\theta)\,(u_1\,u_2)^{-1-\theta}\left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-2-\frac{1}{\theta}}.$$

Example 2.4 (Frank copula)**.** The *Frank copula* with parameter $\theta \in \mathbb{R}\setminus\{0\}$ is an Archimedean copula with generator $\varphi_\theta(t) \coloneqq -\log\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$. We have

$$C(\boldsymbol{u};\theta) = -\frac{1}{\theta}\log\left(1+\frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right), \quad \boldsymbol{u} = (u_1,u_2) \in [0,1]^2, \quad \text{and}$$

$$c(\boldsymbol{u};\theta) = -\frac{\theta\, e^{-\theta(u_1+u_2)}(e^{-\theta}-1)}{\left((e^{-\theta u_1}-1)(e^{-\theta u_2}-1)+e^{-\theta}-1\right)^2}.$$

Example 2.5 (Gumbel copula)**.** The *Gumbel copula* with parameter $\theta \in [1,\infty)$ is an Archimedean copula with generator $\varphi_\theta(t) \coloneqq (-\log t)^\theta$. We obtain

$$C(\boldsymbol{u};\theta) = \exp\left(-\left((-\log u_1)^\theta+(-\log u_2)^\theta\right)^{\frac{1}{\theta}}\right), \quad \boldsymbol{u} = (u_1,u_2) \in [0,1]^2, \quad \text{and}$$

$$c(\boldsymbol{u};\theta) = \frac{C(\boldsymbol{u};\theta)\,(\log u_1 \cdot \log u_2)^{\theta-1}\left(\left((-\log u_1)^\theta + (-\log u_2)^\theta\right)^{\frac{1}{\theta}} + \theta - 1\right)}{u_1\,u_2\left((-\log u_1)^\theta + (-\log u_2)^\theta\right)^{2-\frac{1}{\theta}}}.$$

Density contour plots of above copulas with chosen parameters are displayed in Figure 2.1. These copulas exhibit notable differences in tail behaviour as captured by the lower and upper tail dependence coefficients (TDCs)

$$\lambda_{\mathrm{L}}(C) \coloneqq \lim_{u\to 0}\frac{C(u,u)}{u} \quad \text{and} \quad \lambda_{\mathrm{U}}(C) \coloneqq \lim_{u\to 1}\frac{1-2u+C(u,u)}{1-u}.$$

Moreover, we can express the Kendall's τ of two continuous random variables via their copula using the equality

$$\tau(C) = 4\int_{[0,1]^2} C(u_1,u_2)\,\mathrm{d}C(u_1,u_2) - 1,$$

see Joe (1997, Section 2.1.9). For Archimedean copulas, this equality can equivalently be expressed in terms of the generator φ, yielding

$$\tau(C^\varphi) = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}\,\mathrm{d}t.$$

Table 2.1 summarises the relation between copula parameters, TDCs, and Kendall's τ for the bivariate Clayton, Frank, Gumbel, Gaussian, and Student's copula families.

2.2 Pair-copula constructions (PCCs) and regular vines

While in recent years a vast catalogue of bivariate copula families (also known as *pair-copula* families) has accumulated in the literature, many of these bivariate families have no straightforward multivariate extension. Bedford and Cooke (2001, 2002) introduced a rich and flexible class of multivariate copulas that uses bivariate (conditional) copulas as building blocks only. The corresponding decomposition

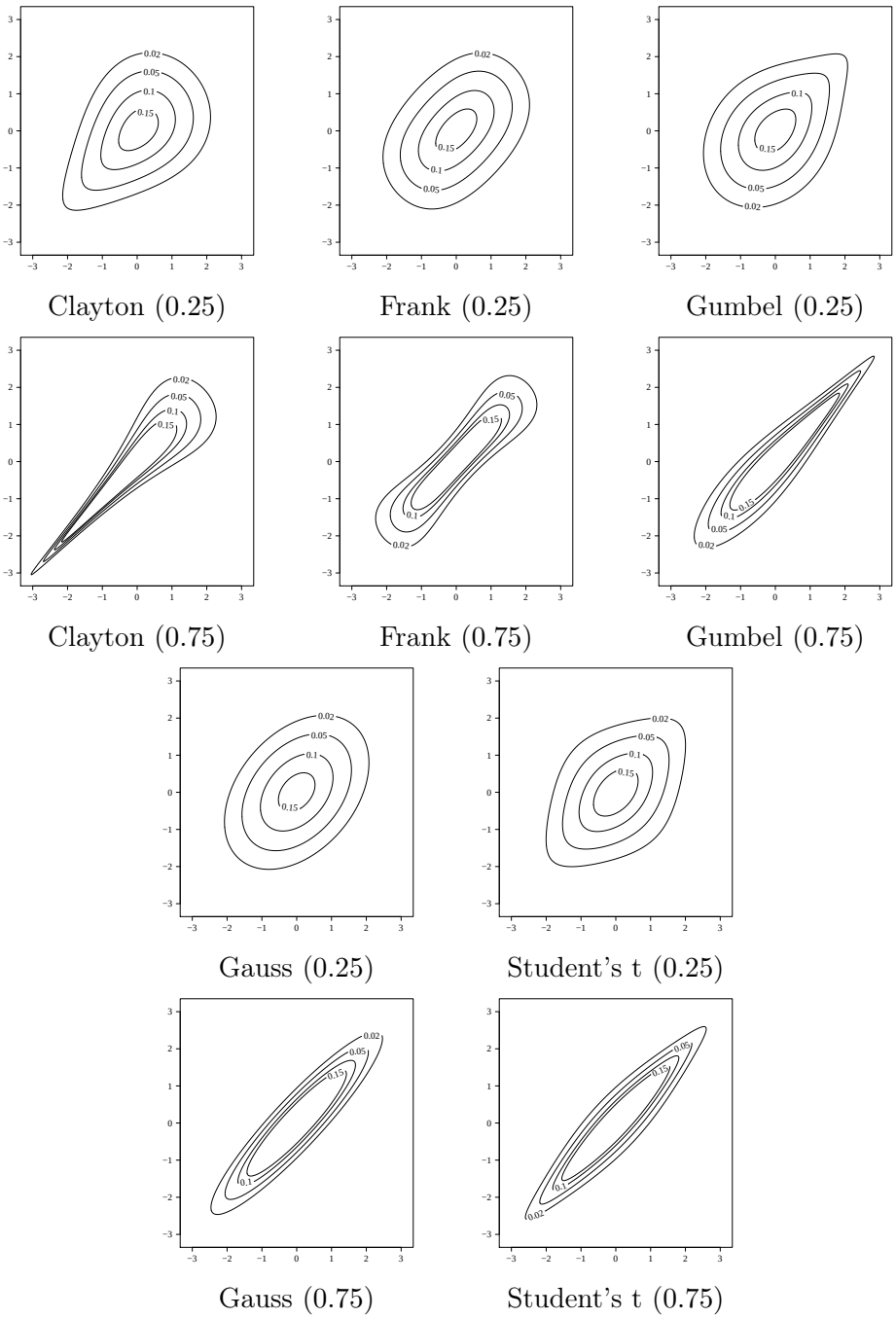


Figure 2.1: Density contour plots of bivariate copulas with Kendall's τ of 0.25 and 0.75 (parentheses). The degrees of freedom of the Student's t copulas were set to 5. All copulas are displayed with standard normal margins.

of a multivariate copula into bivariate copulas is called a *pair-copula construction* (PCC). The most widely researched copulas arising from PCCs are the *vine copulas*. These vine copulas admit a graphical representation called a *regular vine* (R-vine), which consists of a sequence of trees, each edge of which is associated with a certain pair copula in the corresponding PCC. More precisely, let $V \neq \emptyset$ be a finite set and let $d \coloneqq |V|$. An R-vine on V is a sequence $\mathcal{V} \coloneqq (T_1, \dots, T_{d-1})$ of trees $T_1 = (V_1, E_1), \dots, T_{d-1} = (V_{d-1}, E_{d-1})$ such that $V_1 = V$ and $V_i = E_{i-1}$ for $i \geq 2$, that is, the vertices of tree T_i are the edges of tree T_{i-1}. We here represent an edge $v - w$ in tree T_i, $i \in \{1, \dots, d-1\}$, by the doubleton $\{v, w\}$ instead of by the pairs (v, w) and (w, v), that is, $E_i \subseteq \big\{\{v, w\} \,\big|\, v \neq w \in V_i\big\}$. Moreover, every tree T_i, $i \geq 2$, of $\mathcal{V}$ has to satisfy a *proximity condition* requiring that $|v \vartriangle w| = 2$ for every edge $\{v, w\} \in E_i$, where $u \vartriangle v = (u \cup v) \setminus (u \cap v)$. Two vertices in tree T_i, $i \geq 2$, can hence only be adjacent if the corresponding edges in tree T_{i-1} share a common vertex. Last, every edge $\{v, w\} \in E \coloneqq E_1 \cup \dots \cup E_{d-1}$ carries a label $v \vartriangle w \,|\, v \cap w$ representing the (conditional) pair copula $C_{v \vartriangle w | v \cap w}$, where $v \vartriangle w \,|\, \emptyset$ is conveniently replaced by $v \vartriangle w$. Instead of $C_{v \vartriangle w | v \cap w}$ we also write $C_{v_\vartriangle, w_\vartriangle | v \cap w}$, where $v_\vartriangle \coloneqq v \setminus (v \cap w)$ and $w_\vartriangle \coloneqq w \setminus (v \cap w)$. The pdf f of a d-variate probability measure with marginals F_v, $v \in V$, and copula $C_{\mathcal{V}}$ corresponding to $\mathcal{V}$ then takes the form

$$f(\boldsymbol{x}) = \prod_{\{v,w\} \in E} c_{v_\vartriangle, w_\vartriangle | v \cap w}\big(F_{v_\vartriangle | v \cap w}(x_{v_\vartriangle} \,|\, \boldsymbol{x}_{v \cap w}), F_{w_\vartriangle | v \cap w}(x_{w_\vartriangle} \,|\, \boldsymbol{x}_{v \cap w}) \,\big|\, \boldsymbol{x}_{v \cap w}\big) \prod_{v \in V} f_v(x_v), \tag{2.3}$$

where $\boldsymbol{x} = (x_v)_{v \in V} \in \mathbb{R}^d$. Note that—similar to DAGs—the vertices in the first tree of $\mathcal{V}$ represent the univariate margins of $C_{\mathcal{V}}$. In contrast to DAGs, however, $\mathcal{V}$ does not have an interpretation in terms of Markov properties of $C_{\mathcal{V}}$. An example of an R-vine representing a five-variate vine copula is given in Figure 2.2.

For every $K \subseteq V$ and $v \in K$, we define $K_{-v} \coloneqq K \setminus \{v\}$. The conditional cdfs in Equation (2.3) can be evaluated tree-by-tree using a recursive formula derived in Joe (1996), which says that for every $v \in V$, every $K \subseteq V_{-v}$, and an arbitrary $w \in K$

$$F_{v|K}(x_v \,|\, \boldsymbol{x}_K) = \frac{\partial C_{v,w|K_{-w}}\big(F_{v|K_{-w}}(x_v \,|\, \boldsymbol{x}_{K_{-w}}), F_{w|K_{-w}}(x_w \,|\, \boldsymbol{x}_{K_{-w}}) \,\big|\, \boldsymbol{x}_{K_{-w}}\big)}{\partial F_{w|K_{-w}}(x_w \,|\, \boldsymbol{x}_{K_{-w}})}. \tag{2.4}$$

Copula	Kendall's τ	Lower TDC λ_{L}	Upper TDC λ_{U}
Clayton	$\frac{\theta}{\theta+2}$	$2^{-\frac{1}{\theta}}$	0
Frank	$1-\frac{4}{\theta}\big(1-D_1(\theta)\big)$	0	0
Gumbel	$1-\frac{1}{\theta}$	0	$2-2^{\frac{1}{\theta}}$
Gauss	$\frac{2}{\pi}\arcsin(\rho)$	0	0
Student	$\frac{2}{\pi}\arcsin(\rho)$	$2\,\mathrm{t}_{\nu+1}\Big(-\sqrt{\nu+1}\sqrt{\frac{1-\rho}{1+\rho}}\Big)$	$2\,\mathrm{t}_{\nu+1}\Big(-\sqrt{\nu+1}\sqrt{\frac{1-\rho}{1+\rho}}\Big)$

Table 2.1: Relation between parameters, tail dependence coefficients (TDCs), and Kendall's τ for given copula families. D_1 is the first-order Debye function.

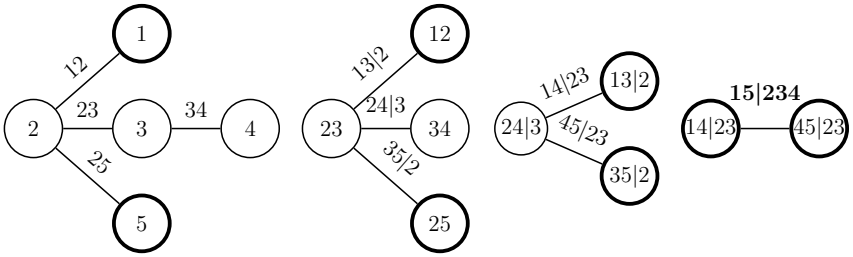


Figure 2.2: An R-vine specifying the pair copulas C_{12}, C_{23}, C_{25}, C_{34}, $C_{13|2}$, $C_{24|3}$, $C_{35|2}$, $C_{14|23}$, $C_{45|23}$, and $C_{15|234}$ (edge labels). Boundaries of vertices including either 1 or 5 appear in bold, see Section 4.1.3.

An iterative algorithm for evaluating the pdf in Equation (2.3) under a simplifying assumption of constant conditional copulas introduced below is given in Dißmann et al. (2013). The first partial derivatives of a pair copula $C_{v,w}$ are also known as *h-functions*. For $(u_v, u_w) \in [0,1]^2$, we write

$$h_{\underline{v},w}(u_v, u_w) \coloneqq \frac{\partial C_{v,w}(u_v, u_w)}{\partial u_w} \quad \text{and} \quad h_{v,\underline{w}}(u_v, u_w) \coloneqq \frac{\partial C_{v,w}(u_v, u_w)}{\partial u_v}.$$

Many popular pair-copula families exhibit closed-form expressions for these h-functions, see Appendix A. Note that by Equation (2.4), we have

$$h_{\underline{v},w}\big(F_v(x_v), F_w(x_w)\big) = F_{v|w}(x_v \,|\, x_w) \quad \text{and} \quad h_{v,\underline{w}}\big(F_v(x_v), F_w(x_w)\big) = F_{w|v}(x_w \,|\, x_v),$$

where $(x_v, x_w) \in \mathbb{R}^2$. Hence, we can extend the notion of h-functions to conditional pair copulas, and express the right hand side of Equation (2.4) by

$$h_{\underline{v},w|K_{-w}}\big(F_{v|K_{-w}}(x_v \,|\, \boldsymbol{x}_{K_{-w}}), F_{w|K_{-w}}(x_w \,|\, \boldsymbol{x}_{K_{-w}}) \,\big|\, \boldsymbol{x}_{K_{-w}}\big).$$

Assume $p \coloneqq |K| \geq 1$, and write $K = \{w_1, \ldots, w_p\}$ such that $w_i \neq w_j$ for $i \neq j$. We define $K_{-i} \coloneqq \{w_{i+1}, \ldots, w_p\}$ for every $i \in \{1, \ldots, p\}$. Observing that $f_{v|K}(x_v \,|\, \boldsymbol{x}_K) = \frac{\mathrm{d}}{\mathrm{d}x_v} F_{v|K}(x_v \,|\, \boldsymbol{x}_K)$, we obtain by the chain rule of differentiation

$$f_{v|K}(x_v|\boldsymbol{x}_K) = f_v(x_v) \prod_{i=1}^{p} c_{v,w_i|K_{-i}}\big(F_{v|K_{-i}}(x_v|\boldsymbol{x}_{K_{-i}}), F_{w_i|K_{-i}}(x_{w_i}|\boldsymbol{x}_{K_{-i}})\big|\boldsymbol{x}_{K_{-i}}\big). \quad (2.5)$$

2.3 ML estimation and model selection in vine copula models

A *vine copula model* is a family of vine copulas together with families of univariate marginals. Maximum likelihood (ML) estimation in vine copula models was first considered in Aas et al. (2009). The findings therein were, however, restricted to vine copula models represented by C- and D-vines. A *C-vine* is an R-vine whose trees are all stars. Conversely, an R-vine is called a *D-vine* if all vertices in tree T_1

are adjacent to at most two other vertices. ML estimation in vine copula models based on general R-vines was considered in Dißmann et al. (2013).

Let $\mathcal{V}$ be an R-vine on V with edge set E, and let $C_{v_\triangle,w_\triangle|v\cap w}(\,\cdot\,,\cdot\,;\boldsymbol{\theta}_{v_\triangle,w_\triangle|v\cap w})$, $\{v,w\}\in E$, be given (conditional) pair copulas with joint parameter vector $\boldsymbol{\theta}:=(\boldsymbol{\theta}_{v_\triangle,w_\triangle|v\cap w})_{\{v,w\}\in E}\in\boldsymbol{\Theta}$. By $\{C_{\mathcal{V},\boldsymbol{\theta}}\,|\,\boldsymbol{\theta}\in\boldsymbol{\Theta}\}$ we denote the corresponding vine copula family. Note that we dropped the values $\boldsymbol{x}_{u\cap v}$ of the conditioning variables from the pair copulas $C_{v_\triangle,w_\triangle|v\cap w}$, thus assuming that the corresponding copula family and parameter vector $\boldsymbol{\theta}_{v_\triangle,w_\triangle|v\cap w}$ remain constant for all $\boldsymbol{x}_{u\cap v}\in\mathbb{R}^{|u\cap v|}$. This *simplifying assumption* is made for computational convenience and has become common practice in likelihood inference for vine copula models, see Hobæk Haff et al. (2010), Acar et al. (2012), and Stöber et al. (2012) for a critical assessment. Furthermore, let $\boldsymbol{u}=(\boldsymbol{u}^1,\dots,\boldsymbol{u}^n)$, $n\in\mathbb{N}$, be a realisation of a sample of i.i.d. observations $\boldsymbol{U}^1,\dots,\boldsymbol{U}^n$ from a random variable $\boldsymbol{U}$ on $[0,1]^d$ with copula family $\{C_{\mathcal{V},\boldsymbol{\theta}}\,|\,\boldsymbol{\theta}\in\boldsymbol{\Theta}\}$ and uniform univariate margins. Equation (2.3) yields the log-likelihood function

$$l(\boldsymbol{\theta};\boldsymbol{u})=\sum_{k=1}^{n}\sum_{\{v,w\}\in E}\log c_{v_\triangle,w_\triangle|v\cap w}\Big(F_{v_\triangle|v\cap w}\big(u^k_{v_\triangle}\,\big|\,\boldsymbol{u}^k_{v\cap w};\boldsymbol{\theta}\big),F_{w_\triangle|v\cap w}\big(u^k_{w_\triangle}\,\big|\,\boldsymbol{u}^k_{v\cap w};\boldsymbol{\theta}\big);\boldsymbol{\theta}\Big). \tag{2.6}$$

The restriction to uniform margins is made for numerical convenience, see below.

2.3.1 ML estimation

Since a joint estimation of the parameters of the univariate marginal distributions and the copula can become computationally demanding in high dimensions, a two-step estimation approach known as the *inference functions for margins* method (Joe and Xu, 1996) is frequently applied. First, the marginal parameters are estimated and second, given the estimates of the marginal parameters, the copula parameters are inferred. In a similar vein, Genest et al. (1995) proposed a semiparametric approach in which the empirical cdf is used to transform the univariate marginals to uniform $[0,1]$ distributions before estimating the parameters of the copula model, see Kim et al. (2007) for a comparison. ML estimation of the parameters

in Equation (2.6) is frequently performed using a stepwise approach as first described in Aas et al. (2009). In a first step, ML estimates of the parameters of each pair-copula family are computed separately. Due to the recursive structure of the log-likelihood function outlined above, this estimation step is carried out tree-by-tree. We refer to the obtained parameter estimates as *sequential ML estimates*. In a second step, the full log-likelihood function is maximised jointly using the sequential ML estimates as starting values, yielding the so-called *joint ML estimates* $\widehat{\boldsymbol{\theta}}_{v_\triangle, w_\triangle | v \cap w}$, $\{v, w\} \in E$. Large and small sample applications of the stepwise estimation procedure have shown that the sequential ML estimates also provide a good approximation of their joint counterparts, see Hobæk Haff (2013, 2012) for consistency results and a simulation study. One might hence consider omitting the second estimation step in a given situation to reduce computational complexity.

2.3.2 Model selection

Model selection for vine copula models comprises estimation of the R-vine $\mathcal{V}$ and selection of the pair-copula families for $C_{v_\triangle, w_\triangle | v \cap w}$, $\{v, w\} \in E$. Given $\mathcal{V}$, the latter task of selecting pair-copula families can be performed tree-by-tree, choosing for each edge $\{v, w\} \in E$ the one pair-copula family among a given set of candidate families that optimises a given selection criterion like Akaike's information criterion (AIC) (Akaike, 1974) or the Bayesian information criterion (BIC) (Schwarz, 1978). Dißmann et al. (2013) presented a greedy-type algorithm for the estimation of $\mathcal{V}$, which estimates the trees $T_1, \dots, T_{d-2}$ sequentially, that is, again tree-by-tree. Note that estimating tree T_{d-2} also fixes tree T_{d-1}. Structure estimation for tree $T_i = (V_i, E_i)$, $i \in \{1, \dots, d-2\}$, is carried out in three steps. In a first step, a weight $\omega_{v,w}$ is assigned to every pair of vertices $v, w \in V_i$ with $|v \triangle w| = 2$. Suitable weights given the data are, for instance, the absolute values of estimates of Kendall's τ, or AIC or BIC values of selected pair-copula families with estimated parameters. In a second step, T_i is set to be a tree on V_i optimising the sum of edge weights $\sum_{\{v,w\} \in E_i} \omega_{v,w}$, where $|v \triangle w| = 2$ for all $\{v, w\} \in E_i$ to ensure the proximity condition. Such an *optimal spanning tree* can be found using the algorithms by Kruskal (1956) or Prim (1957). In a last step, a pair-copula family is assigned to each edge $\{v, w\} \in E_i$,

as described above, and an ML estimate of the corresponding parameter(s) is computed. This last step may have already been performed when computing the edge weights $\omega_{v,w}$. Note that due to the greedy nature of the algorithm, the resulting R-vine need not optimise the sum of all edge weights $\sum_{\{v,w\}\in E} \omega_{v,w}$. The search for optimal spanning trees reduces to a search for root vertices when only considering C-vines instead of the more general R-vines, cf. Czado et al. (2012). Since a D-vine is completely determined by tree T_1, only one tree has to be specified when restricting the class of R-vines to D-vines. Due to the particular structure of D-vines, however, finding tree T_1 by the above method leads to a *travelling salesman problem* (TSP) (Applegate et al., 2007), which is NP-hard. Kurowicka (2011) proposed an alternative structure selection algorithm, in which $\mathcal{V}$ is built in reverse order from tree T_{d-1} to tree T_1 using partial correlation estimates as weights. Bayesian approaches to structure estimation have been considered in Smith et al. (2010), Min and Czado (2011), and Gruber et al. (2013). Implementations of model selection and ML estimation procedures for vine copula models are readily available in the R package `VineCopula` (Schepsmeier et al., 2013).

The construction of a d-variate vine copula model requires the specification of $\binom{d}{2}$ pair-copula families, a number growing quadratically in d. The actual number of decisions to make in practical applications may, however, be lower if we happen to discover (conditional) independences in the analysed data. In that case, the corresponding pair copulas are set to be independence copulas. Since above structure estimation algorithm is based on the idea of modelling strongest dependences in the first trees, Brechmann et al. (2012) proposed to set all pair copulas in the later trees to independence copulas, which leads to so-called *truncated R-vines*. Instead of leaving the detection of (conditional) independences to chance, one may, however, consider modelling these independences in the first place to obtain more parsimonious models. Unfortunately, the construction of vine copula models satisfying prespecified conditional independence restrictions is a hard problem in general. A class of models suited for this task are the Bayesian networks discussed in Chapter 1. Kurowicka and Cooke (2005) hence joined graphical and copula modelling to introduce PCCs for Bayesian networks, which we will investigate in the next chapter.

3 Pair-copula Bayesian networks

Pair-copula Bayesian networks (PCBNs) combine the distributional flexibility of PCCs with the parsimony of Bayesian networks. In this chapter, we derive PCCs for $\mathcal{D}$-Markovian probability measures and develop routines for computing conditional cdfs. Moreover, we consider random sampling in PCBNs and investigate these models' aptitude for likelihood inference in a simulation study. PCBNs were first introduced in Kurowicka and Cooke (2005) and further extended in Hanea (2008). However, the analyses therein were restricted to pair-copula families with the property that zero rank correlation implies independence. Also, these authors concentrate on non-parametric statistical inference, while we focus attention to parametric likelihood inference.

3.1 PCCs for D-Markovian probability measures

Let $\mathcal{D} = (V, E)$ be a DAG, and let P be an absolutely continuous $\mathcal{D}$-Markovian probability measure on $\mathbb{R}^d$, $d \coloneqq |V|$, with strictly increasing univariate marginal cdfs. We aim to derive a pair-copula decomposition for the pdf f of P. To this end, we order the parent sets of all vertices in $\mathcal{D}$ increasingly. Let $w_v \colon \{1, \dots, |\mathrm{pa}(v)|\} \to \mathrm{pa}(v)$, $i \mapsto w_i \coloneqq w_v(i)$, be a bijection for every $v \in V$ with $|\mathrm{pa}(v)| \geq 1$. We introduce a total order $<_v$ on $\mathrm{pa}(v)$ for every $v \in V$ such that whenever $|\mathrm{pa}(v)| \geq 1$ we have $w_i <_v w_j$ if and only if $i < j$ for all $i, j \in \{1, \dots, |\mathrm{pa}(v)|\}$. Note that there are $|\mathrm{pa}(v)|!$ permutations of $\mathrm{pa}(v)$ (up to isomorphism). We call $\mathcal{O} \coloneqq \{<_v \mid v \in V\}$ a set of *parent orderings* for $\mathcal{D}$. For every $v \in V$ and $w \in \mathrm{pa}(v)$, we set

$$\mathrm{pa}(v; w) \coloneqq \{u \in \mathrm{pa}(v) \mid u <_v w\} = \{w_i \in \mathrm{pa}(v) \mid i < w_v^{-1}(w)\}.$$

By Sklar's theorem, we know that the cdf of P can be uniquely decomposed into the univariate marginals $F_1, \dots, F_d$ and a copula C. The following theorem shows that C can be further decomposed into the (conditional) pair copulas $C_{v,w|\mathrm{pa}(v;w)}$, $v \in V$, $w \in \mathrm{pa}(v)$, which yields a PCC for C in which each (conditional) pair copula corresponds to exactly one edge $w \to v$ in $\mathcal{D}$. Contrary to the vine-based approach of Kurowicka and Cooke (2005), the proof relies on graph-theoretical considerations only.

Theorem 3.1. *Let $\mathcal{D} = (V, E)$ be a DAG on $d \coloneqq |V|$ vertices, and let P be an absolutely continuous $\mathcal{D}$-Markovian probability measure on $\mathbb{R}^d$ with strictly increasing univariate marginal cdfs. Then P is uniquely determined by its univariate margins P_v, $v \in V$, and its (conditional) pair copulas $C_{v,w|\mathrm{pa}(v;w)}$, $v \in V$, $w \in \mathrm{pa}(v)$.*

Proof. (Induction on the cardinality of V) Since P is $\mathcal{D}$-Markovian, its pdf f admits a $\mathcal{D}$-recursive factorisation of the form (1.3). The claim is trivial for $|V| = 1$. Hence, let $|V| \geq 2$. Since $\mathcal{D}$ is acyclic, we may choose some maximal vertex of $\mathcal{D}$, that is, some $m \in V$ with $\mathrm{de}(m) = \emptyset$. Let $V' \coloneqq V \setminus \{m\}$ and $E' \coloneqq E \cap (V' \times V')$. Then above-mentioned factorisation can be written as

$$f(\boldsymbol{x}) = f_{m|\mathrm{pa}(m)}(x_m \,|\, \boldsymbol{x}_{\mathrm{pa}(m)}) \prod_{v \in V'} f_{v|\mathrm{pa}(v)}(x_v \,|\, \boldsymbol{x}_{\mathrm{pa}(v)}), \quad \boldsymbol{x} = (x_v)_{v \in V} \in \mathbb{R}^d. \tag{3.1}$$

By the choice of m, the sets $\mathrm{pa}(v)$ and $\mathrm{nd}(v)$, $v \in V'$, remain unaffected by a transition from $\mathcal{D}$ to the subgraph $\mathcal{D}' = (V', E')$. Thus, $P_{V'}$ is $\mathcal{D}'$-Markovian and the product $\prod_{v \in V'} f_{v|\mathrm{pa}(v)}(...)$ on the right hand side of Equation (3.1) is the $\mathcal{D}'$-recursive factorisation of $f_{V'}$. We may assume inductively that $P_{V'}$, and hence $f_{V'}$, is uniquely determined by the univariate margins P_v, $v \in V'$, and the (conditional) pair copulas $C_{v,w|\mathrm{pa}(v;w)}$, $v \in V'$, $w \in \mathrm{pa}(v)$. It remains to show that the same property holds for $f_{m|\mathrm{pa}(m)}$ if we include P_m and $C_{m,w|\mathrm{pa}(m;w)}$, $w \in \mathrm{pa}(m)$, in our analysis.

We prove the latter claim by induction on $k \coloneqq |\mathrm{pa}(m)|$. The claim is trivial for $k = 0$. For $k \geq 1$, let $w_1 <_m \dots <_m w_k$ denote the elements of $\mathrm{pa}(m)$ and let $W \coloneqq \mathrm{pa}(m; w_k)$. By Equation (2.4),

$$f_{m|\mathrm{pa}(m)}(x_m \,|\, \boldsymbol{x}_{\mathrm{pa}(m)}) = c_{m,w_k|W}\big(F_{m|W}(x_m \,|\, \boldsymbol{x}_W), F_{w_k|W}(x_{w_k} \,|\, \boldsymbol{x}_W) \,\big|\, \boldsymbol{x}_W\big)\, f_{m|W}(x_m \,|\, \boldsymbol{x}_W).$$

Since $\mathrm{pa}(m) \subseteq V'$, the conditional cdf $F_{w_k|W}(\,\cdot \,|\, \boldsymbol{x}_W)$ is completely determined by $P_{V'}$ and therefore by the quantities specified in the theorem's claim. Observing that $W = \mathrm{pa}(m) \setminus \{w_k\}$, we may conclude by induction that $F_{m|W}(\,\cdot \,|\, \boldsymbol{x}_W)$ and $f_{m|W}(\,\cdot \,|\, \boldsymbol{x}_W)$ also exhibit the claimed property. This establishes the claim. □

By the above theorem, we can write $f(\boldsymbol{x})$, $\boldsymbol{x} = (x_v)_{v\in V} \in \mathbb{R}^d$, as

$$\prod_{v\in V} f_v(x_v) \prod_{w\in \mathrm{pa}(v)} c_{v,w|\mathrm{pa}(v;w)}\big(F_{v|\mathrm{pa}(v;w)}(x_v|\boldsymbol{x}_{\mathrm{pa}(v;w)}), F_{w|\mathrm{pa}(v;w)}(x_w|\boldsymbol{x}_{\mathrm{pa}(v;w)})\big|\boldsymbol{x}_{\mathrm{pa}(v;w)}\big). \quad (3.2)$$

We may again distinguish the pair copulas involved by the number of conditioning variables. Similar to vine copulas, this number ranges between 0 and possibly $d-1$. Although the graphical representations of the two models look fairly similar, the concepts behind are completely different. While R-vines illustrate the required pair copulas only, the edges of a DAG specify conditional independence restrictions. In both cases, however, these representations are visual aids only and can be omitted.

The remaining question is whether DAGs actually yield pair-copula decompositions which are not representable by R-vines. This question is closely related to the computation of the conditional cdfs $F_{v|\mathrm{pa}(v;w)}(\,\cdot \,|\, \boldsymbol{x}_{\mathrm{pa}(v;w)})$ and $F_{w|\mathrm{pa}(v;w)}(\,\cdot \,|\, \boldsymbol{x}_{\mathrm{pa}(v;w)})$ in Equation (3.2). The answer is yes. In contrast to R-vines, DAGs allow the specification of conditional cdfs that cannot be computed by simply plugging in results from preceding trees. Hence, these conditional cdfs have to be computed via integration of other margins. In view of their application for statistical inference, however, DAG PCCs are more parsimonious than R-vine PCCs, see Section 3.4.

Example 3.1. Consider the DAG $\mathcal{D}$ in Figure 1.2 with ordering $2 <_4 3$ of $\mathrm{pa}(4) = \{2,3\}$. By Table 1.1, we obtain $\mathrm{pa}(1;\emptyset) = \emptyset$, $\mathrm{pa}(2;1) = \emptyset$, $\mathrm{pa}(3;1) = \emptyset$, $\mathrm{pa}(4;2) = \emptyset$, and $\mathrm{pa}(4;3) = \{2\}$. Equation (3.2) yields

$$\begin{aligned} f(\boldsymbol{x}) = f_1(x_1)\cdots f_4(x_4) &\cdot c_{21}\big(F_2(x_2), F_1(x_1)\big) \cdot c_{31}\big(F_3(x_3), F_1(x_1)\big) \cdot c_{42}\big(F_4(x_4), F_2(x_2)\big) \\ &\cdot c_{43|2}\big(F_{4|2}(x_4\,|\,x_2), F_{3|2}(x_3\,|\,x_2)\,\big|\,x_2\big), \quad \boldsymbol{x} = (x_1,\ldots,x_4) \in \mathbb{R}^4, \end{aligned}$$

where $F_{4|2}(x_4\,|\,x_2) = h_{\underline{4}2}\big(F_4(x_4), F_2(x_2)\big)$ by Equation (2.4). Since the copula C_{32}

is not available in the decomposition of f, we exploit the conditional independence $X_2 \perp\!\!\!\perp X_3 \mid X_1$ to obtain

$$\begin{aligned}
F_{3|2}(x_3 \,|\, x_2) &= \int_{-\infty}^{x_3} f_{3|2}(y_3 \,|\, x_2)\,\mathrm{d}y_3 = f_2^{-1}(x_2) \int_{-\infty}^{x_3} f_{23}(x_2, y_3)\,\mathrm{d}y_3 \\
&= f_2^{-1}(x_2) \int_{-\infty}^{x_3} \int_{-\infty}^{\infty} f_{123}(y_1, x_2, y_3)\,\mathrm{d}y_1\,\mathrm{d}y_3 \\
&\overset{(1.3)}{=} f_2^{-1}(x_2) \int_{-\infty}^{\infty} f_1(y_1)\, f_{2|1}(x_2 \,|\, y_1) \int_{-\infty}^{x_3} f_{3|1}(y_3 \,|\, y_1)\,\mathrm{d}y_3\,\mathrm{d}y_1 \\
&= \int_{-\infty}^{\infty} f_1^{-1}(y_1)\, f_2^{-1}(x_2)\, f_{12}(y_1, x_2)\, F_{3|1}(x_3 \,|\, y_1)\,\mathrm{d}F_1(y_1) \\
&\overset{(2.4)}{=} \int_{-\infty}^{\infty} c_{21}\big(F_2(x_2), F_1(y_1)\big)\, \frac{\partial C_{31}\big(F_3(x_3), F_1(y_1)\big)}{\partial F_1(y_1)}\,\mathrm{d}F_1(y_1) \\
&= \int_0^1 c_{21}\big(F_2(x_2), u_1\big)\, \frac{\partial C_{31}\big(F_3(x_3), u_1\big)}{\partial u_1}\,\mathrm{d}u_1 \\
&= \int_0^1 c_{21}\big(F_2(x_2), u_1\big)\, h_{\underline{3}1}\big(F_3(x_3), u_1\big)\,\mathrm{d}u_1.
\end{aligned}$$

There is no general closed-form solution for the last integral. If we instead choose the ordering $3 <_4 2$ for $\mathrm{pa}(4) = \{2, 3\}$, we obtain the same decomposition as above with the roles of vertices 2 and 3 interchanged.

Due to the appearing integral, the pair-copula decomposition in the example cannot be represented by an R-vine. There are, however, DAG PCCs representable by R-vines, see Section 3.4 for a four-variate DAG PCC that coincides with a D-vine PCC. Note that straightforward application of Sklar's theorem to Equation (1.3) yields another copula decomposition for f, see Elidan (2010). This decomposition, however, consists of generally higher-variate copulas and hence leads to statistical models hampered by the difficulties described in Section 2.2.

3.2 Evaluating conditional cdfs

As shown above, the challenge in Equation (3.2) lies in the evaluation of the conditional cdfs. Assume without loss of generality that $\mathcal{D}$ is well-ordered. Let $v \in V$

and let $J \subseteq V \setminus \{v\}$ be non-empty. We will now derive a pair-copula decomposition for the conditional cdf $F_{v|J}(\,\cdot \,|\, \boldsymbol{x}_J)$. We begin by exploiting the (conditional) independence restrictions represented by $\mathcal{D}$. To this end, consider the moral graph $\mathcal{G} \coloneqq (\mathcal{D}_{\mathrm{An}(\{v\}\cup J)})^m$. If $\{v\} \perp I \mid (J \setminus I)\,[\mathcal{G}]$ for some non-empty $I \subseteq J$, then the global $\mathcal{D}$-Markov property in Equation (1.2) yields with $K \coloneqq J \setminus I$

$$f_{v|J}(x_v \,|\, \boldsymbol{x}_J) = \frac{f_{\{v\}\cup J}(\boldsymbol{x}_{\{v\}\cup J})}{f_J(\boldsymbol{x}_J)} = \frac{f_{v|K}(x_v \,|\, \boldsymbol{x}_K)\, f_{I|K}(\boldsymbol{x}_I \,|\, \boldsymbol{x}_K)\, f_K(\boldsymbol{x}_K)}{f_{I|K}(\boldsymbol{x}_I \,|\, \boldsymbol{x}_K)\, f_K(\boldsymbol{x}_K)} = f_{v|K}(x_v \,|\, \boldsymbol{x}_K),$$

where by convention $f_{W|\emptyset}(\boldsymbol{x}_W \,|\, \boldsymbol{x}_\emptyset) \coloneqq f_W(\boldsymbol{x}_W)$ for every $W \subseteq V$, and $f_\emptyset(\boldsymbol{x}_\emptyset) \coloneqq 1$. Thus, $F_{v|J}(\,\cdot \,|\, \boldsymbol{x}_J) = F_{v|K}(\,\cdot \,|\, \boldsymbol{x}_K)$, and we can continue with the conditioning set K. The case $K = \emptyset$ is trivial. Assume $K \neq \emptyset$. Observing that

$$F_{v|K}(y \,|\, \boldsymbol{x}_K) = \frac{\int_{-\infty}^{y} f_{\{v\}\cup K}(\boldsymbol{x}_{\{v\}\cup K})\,\mathrm{d}x_v}{f_K(\boldsymbol{x}_K)}, \tag{3.3}$$

we next need to find pair-copula decompositions for $f_{\{v\}\cup K}$ and f_K.

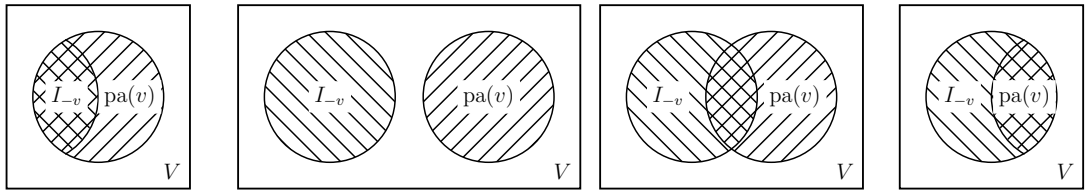
3.2.1 Pair-copula decompositions for marginal pdfs

More generally, let $I \subseteq V$ be non-empty and consider the (marginal) pdf f_I. For every $v \in I$, we set $I_{-v} \coloneqq I \setminus \{v\}$ and obtain the following lemma.

Lemma 3.2. *Let $\mathcal{D} = (V, E)$ be a well-ordered DAG on $d \coloneqq |V|$ vertices, and let P be an absolutely continuous $\mathcal{D}$-Markovian probability measure on $\mathbb{R}^d$ with pdf f. Let $I \subseteq V$ be non-empty and let v denote the maximal vertex in I by the well-ordering of $\mathcal{D}$. Moreover, define $S_v \coloneqq \big\{u \in \mathrm{pa}(v) \,\big|\, \{u\} \perp I_{-v}\,[(\mathcal{D}_{\mathrm{An}(\{u\}\cup I_{-v})})^m]\big\}$ and*

$$W_v \coloneqq \begin{cases} \emptyset & \text{if } I_{-v} = \emptyset \text{ or } S_v = \mathrm{pa}(v), \\ \{w_1\} \cup \mathrm{pa}(v; w_1) & \text{if } I_{-v} \subseteq \mathrm{pa}(v) \text{ and } I_{-v} \neq \emptyset, \\ \{w_2\} \cup \mathrm{pa}(v; w_2) & \text{else,} \end{cases}$$

where w_1 and w_2 denote the maximal vertices in I_{-v} and $\mathrm{pa}(v) \setminus S_v$, respectively, by



(a) $W_v = \{w_1\} \cup \mathrm{pa}(v; w_1)$ (b) $W_v = \emptyset$ or $W_v = \{w_2\} \cup \mathrm{pa}(v; w_2)$ (c) $W_v = \{w_2\} \cup \mathrm{pa}(v; w_2)$ (d) $W_v = \mathrm{pa}(v)$

Figure 3.1: Venn diagrams of the sets $I_{-v} \neq \emptyset$ and $\mathrm{pa}(v) \neq \emptyset$, and corresponding definitions of W_v (see lower captions).

a given parent ordering $<_v$*. Then for all* $\boldsymbol{x}_I = (x_u)_{u \in I} \in R^{|I|}$*,*

$$f_I(\boldsymbol{x}_I) = \int_{\mathbb{R}^{|W_v \setminus I|}} f_{v|W_v}(x_v \mid \boldsymbol{x}_{W_v})\, f_{W_v \cup I_{-v}}(\boldsymbol{x}_{W_v \cup I_{-v}})\, \mathrm{d}\boldsymbol{x}_{W_v \setminus I}. \tag{3.4}$$

Note that by convention, $\int_{\mathbb{R}^0} g(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}_\emptyset := g(\boldsymbol{x})$ for every integrable function $g\colon \mathbb{R}^k \to \mathbb{R}$, $k \in \mathbb{N}$. Also, the parent ordering $<_v$ need not concur with the well-ordering of $\mathcal{D}$.

Proof. As can be seen from the definition of W_v, the decomposition of f_I in the lemma's claim depends on the relation between the sets I_{-v} and $\mathrm{pa}(v)$. Assume first that $I_{-v} = \emptyset$. Then $f_I = f_v$ and the claim is trivial.

Next, assume $I_{-v} \neq \emptyset$ but $\mathrm{pa}(v) = \emptyset$. Then $S_v = \emptyset$. Since v is maximal in I by the well-ordering of $\mathcal{D}$, v has no descendants in I_{-v}, and we have $\{v\} \perp I_{-v}\, [(\mathcal{D}_{\mathrm{An}(I)})^m]$. The global $\mathcal{D}$-Markov property thus yields $f_I(\boldsymbol{x}_I) = f_v(x_v)\, f_{I_{-v}}(\boldsymbol{x}_{I_{-v}})$, that is, Equation (3.4) for $W_v := \emptyset$.

From now on assume $I_{-v} \neq \emptyset$ and $\mathrm{pa}(v) \neq \emptyset$. The possible relations between I_{-v} and $\mathrm{pa}(v)$ are illustrated in Figure 3.1. If $I_{-v} \subseteq \mathrm{pa}(v)$ (Figure 3.1a), we extend I_{-v} to $W_v := \{w_1\} \cup \mathrm{pa}(v; w_1) \supseteq I_{-v}$ and obtain as claimed

$$f_I(\boldsymbol{x}_I) = \int_{\mathbb{R}^{|W_v \setminus I|}} f_{\{v\} \cup W_v}(\boldsymbol{x}_{\{v\} \cup W_v})\, \mathrm{d}\boldsymbol{x}_{W_v \setminus I} = \int_{\mathbb{R}^{|W_v \setminus I|}} f_{v|W_v}(x_v \mid \boldsymbol{x}_{W_v})\, f_{W_v}(\boldsymbol{x}_{W_v})\, \mathrm{d}\boldsymbol{x}_{W_v \setminus I}.$$

Note that in case $I_{-v} = \{w_1\} \cup \mathrm{pa}(v; w_1)$, no integration is required since $W_v \setminus I = \emptyset$.

Next, let $I_{-v} \cap \mathrm{pa}(v) = \emptyset$ (Figure 3.1b). If $S_v = \mathrm{pa}(v)$, then $\{v\} \perp I_{-v}\, [(\mathcal{D}_{\mathrm{An}(I)})^m]$ since v has no descendants in I_{-v}. Hence, we again have $f_I(\boldsymbol{x}_I) = f_v(x_v)\, f_{I_{-v}}(\boldsymbol{x}_{I_{-v}})$,

that is, Equation (3.4) for $W_v \coloneqq \emptyset$. If, however, $S_v \neq \mathrm{pa}(v)$, then $\{v\} \perp I_{-v} \mid W_v\,[(\mathcal{D}_{\mathrm{An}(I\cup W_v)})^m]$, where $W_v \coloneqq \{w_2\} \cup \mathrm{pa}(v; w_2)$, and the global $\mathcal{D}$-Markov property yields

$$f_{I\cup W_v}(\boldsymbol{x}_{I\cup W_v}) = f_{v|W_v}(x_v \,|\, \boldsymbol{x}_{W_v})\, f_{I_{-v}|W_v}(\boldsymbol{x}_{I_{-v}} \,|\, \boldsymbol{x}_{W_v})\, f_{W_v}(\boldsymbol{x}_{W_v}). \tag{3.5}$$

Since $I_{-v} \cap W_v = \emptyset$, we thus get

$$f_I(\boldsymbol{x}_I) = \int_{\mathbb{R}^{|W_v|}} f_{v|W_v}(x_v \,|\, \boldsymbol{x}_{W_v})\, f_{W_v\cup I_{-v}}(\boldsymbol{x}_{W_v\cup I_{-v}})\,\mathrm{d}\boldsymbol{x}_{W_v}.$$

Note that in case $S_v = \emptyset$, we have $W_v = \mathrm{pa}(v)$.

Finally, assume $I_{-v} \cap \mathrm{pa}(v) \neq \emptyset$ such that $I_{-v} \nsubseteq \mathrm{pa}(v)$ (Figure 3.1c). Similarly to Equation (3.5), we obtain with $W_v \coloneqq \{w_2\} \cup \mathrm{pa}(v; w_2)$

$$f_{I\cup W_v}(\boldsymbol{x}_{I\cup W_v}) = f_{v|W_v}(x_v \,|\, \boldsymbol{x}_{W_v})\, f_{(I_{-v}\setminus W_v)|W_v}(\boldsymbol{x}_{I_{-v}\setminus W_v} \,|\, \boldsymbol{x}_{W_v})\, f_{W_v}(\boldsymbol{x}_{W_v})$$

by the global $\mathcal{D}$-Markov property, and hence

$$f_I(\boldsymbol{x}_I) = \int_{\mathbb{R}^{|W_v\setminus I|}} f_{v|W_v}(x_v \,|\, \boldsymbol{x}_{W_v})\, f_{W_v\cup I_{-v}}(\boldsymbol{x}_{W_v\cup I_{-v}})\,\mathrm{d}\boldsymbol{x}_{W_v\setminus I}.$$

Note again that in case $S_v = \emptyset$, we have $W_v = \mathrm{pa}(v)$. Also, note that in case $\mathrm{pa}(v) \subseteq I_{-v}$ (Figure 3.1d), no integration is required. This establishes the claim. □

The set W_v in Lemma 3.2 is either empty or of the form $\{w\} \cup \mathrm{pa}(v; w)$ for some $w \in \mathrm{pa}(v)$. In the latter case, we can express $f_{v|\{w\}\cup\mathrm{pa}(v;w)}(\,\cdot \,|\, \boldsymbol{x}_{\{w\}\cup\mathrm{pa}(v;w)})$ on the right hand side of Equation (3.4) in terms of the univariate marginals F_u, $u \in V$, and the (conditional) pair copulas $C_{v,u|\,\mathrm{pa}(v;u)}$, $u \in \mathrm{pa}(v)$, as follows.

Lemma 3.3. *Let the notation be as in Lemma 3.2 and let P have strictly increasing univariate marginal cdfs. Let $I_{-v} \neq \emptyset$ and let $S_v \neq \mathrm{pa}(v)$. Then $f_{v|W_v}(x_v \,|\, \boldsymbol{x}_{W_v})$, $x_v \in \mathbb{R}$, $\boldsymbol{x}_{W_v} = (x_w)_{w\in W_v} \in \mathbb{R}^{|W_v|}$, takes the form*

$$f_v(x_v) \prod_{w\in W_v} c_{v,w|\mathrm{pa}(v;w)}\big(F_{v|\mathrm{pa}(v;w)}(x_v \,|\, \boldsymbol{x}_{\mathrm{pa}(v;w)}), F_{w|\mathrm{pa}(v;w)}(x_w \,|\, \boldsymbol{x}_{\mathrm{pa}(v;w)}) \,\big|\, \boldsymbol{x}_{\mathrm{pa}(v;w)}\big).$$

Proof. Since $I_{-v} \neq \emptyset$ and $S_v \neq \mathrm{pa}(v)$, W_v is non-empty and thus $W_v = \{u\} \cup \mathrm{pa}(v;u)$ for some $u \in \mathrm{pa}(v)$. By Equation (2.5), we can hence write $f_{v|W_v}(x_v \,|\, \boldsymbol{x}_{W_v})$ as

$$f_v(x_v) \prod_{w \in W_v} c_{v,w|\mathrm{pa}(v;w)}\big(F_{v|\mathrm{pa}(v;w)}(x_v \,|\, \boldsymbol{x}_{\mathrm{pa}(v;w)}), F_{w|\mathrm{pa}(v;w)}(x_w \,|\, \boldsymbol{x}_{\mathrm{pa}(v;w)}) \,\big|\, \boldsymbol{x}_{\mathrm{pa}(v;w)}\big),$$

and the claim is proven. □

Since all vertices in $W_v \cup I_{-v}$ are smaller than v by the well-ordering of $\mathcal{D}$ and since V is finite, we can inductively apply Lemmas 3.2 and 3.3 to the pdf $f_{W_v \cup I_{-v}}$ in Equation (3.4) until no unconditional pdfs of dimension higher than one remain. Let J denote the set of vertices corresponding to the integration variables added during this iterative procedure (and including $W_v \setminus I$). Given a set $\mathcal{O}$ of parent orderings for $\mathcal{D}$, Lemma 3.2 yields a set W_u for every $u \in I \cup J$. We have hence established the following theorem.

Theorem 3.4. *Let the notation be as in Lemma 3.3. Then*

$$\begin{aligned} f_I(\boldsymbol{x}_I) = \int_{\mathbb{R}^{|J|}} & \prod_{v \in (I \cup J)} f_v(x_v) \\ & \times \prod_{w \in W_v} c_{v,w|\mathrm{pa}(v;w)}\big(F_{v|\mathrm{pa}(v;w)}(x_v \,|\, \boldsymbol{x}_{\mathrm{pa}(v;w)}), F_{w|\mathrm{pa}(v;w)}(x_w \,|\, \boldsymbol{x}_{\mathrm{pa}(v;w)}) \,\big|\, \boldsymbol{x}_{\mathrm{pa}(v;w)}\big)\, \mathrm{d}\boldsymbol{x}_J \end{aligned}$$

for all $\boldsymbol{x}_I = (x_v)_{v \in I} \in R^{|I|}$. □

Note that in the special case $I = V$, Theorem 3.4 yields Equation (3.2). Above procedure for deriving a pair-copula decomposition of f_I as given in Theorem 3.4 is summarised in Algorithm 1.

Example 3.2. Consider the well-ordered DAG $\mathcal{G}$ in Figure 3.2. The edges and parent orderings of $\mathcal{G}$ can be summarised in a matrix $A_{\mathcal{G}} = (a_{ij})_{1 \leq i,j \leq 7}$ whose elements satisfy $a_{ij} = k$, $k \leq |\mathrm{pa}(j)|$, if $\mathcal{G}$ contains the edge $i \to j$ and if i is the k-th smallest parent of j by $<_j$, and $a_{ij} = 0$ otherwise, see Figure 3.2. For the reader's

Algorithm 1 Pair-copula decomposition of a (marginal) pdf.

Input Well-ordered DAG $\mathcal{D}$; set of parent orderings $\mathcal{O}$; non-empty vertex set $I \subseteq V$.
Output Factorisation f. % (marginal) pdf $f_I(\boldsymbol{x}_I)$

$f \leftarrow 1$;
$J \leftarrow \emptyset$; % indices of integration variables
while $|I| \geq 1$ **do**
  % Select maximal vertex:
  $v \leftarrow$ maximal vertex in I by the well-ordering of $\mathcal{D}$;
  $f \leftarrow f \cdot f_v(x_v)$;
  $I \leftarrow I_{-v}$;
  % Determine the set W_v:
  $W \leftarrow \emptyset$;
  $S \leftarrow \big\{ w \in \mathrm{pa}(v) \,\big|\, \{w\} \perp I \,[(\mathcal{D}_{\mathrm{An}(\{w\} \cup I)})^m] \big\}$;
  if $I \neq \emptyset$ **and** $S \neq \mathrm{pa}(v)$ **then**
    if $I \subseteq \mathrm{pa}(v)$ **then**
      $w \leftarrow$ maximal vertex in I by the parent ordering $<_v$;
      $W \leftarrow \{w\} \cup \mathrm{pa}(v; w)$;
    else
      $w \leftarrow$ maximal vertex in $\mathrm{pa}(v) \setminus S$ by the parent ordering $<_v$;
      $W \leftarrow \{w\} \cup \mathrm{pa}(v; w)$;
    end if
  end if
  % Introduce corresponding pair copulas and integration variables:
  for $w \in W$ **do**
    $f \leftarrow f \cdot c_{v,w|\mathrm{pa}(v;w)}\big(F_{v|\mathrm{pa}(v;w)}(x_v|\boldsymbol{x}_{\mathrm{pa}(v;w)}), F_{w|\mathrm{pa}(v;w)}(x_w|\boldsymbol{x}_{\mathrm{pa}(v;w)}) \big| \boldsymbol{x}_{\mathrm{pa}(v;w)}\big)$;
    if $w \notin I$ **then**
      $I \leftarrow I \cup \{w\}$;
      $J \leftarrow J \cup \{w\}$;
    end if
  end for
end while
$f \leftarrow \int_{\mathbb{R}^{|J|}} f \,\mathrm{d}\boldsymbol{x}_J$;

convenience we omit function arguments. Equation (3.2) yields

$$\begin{aligned} f = f_1 \cdots f_7 \, & c_{21}(F_2, F_1) \, c_{31}(F_3, F_1) \, c_{42}(F_4, F_2) \, c_{41|2}(F_{4|2}, F_{1|2}) \, c_{54}(F_5, F_4) \, c_{53|4}(F_{5|4}, F_{3|4}) \\ & \cdot c_{65}(F_6, F_5) \, c_{64|5}(F_{6|5}, F_{4|5}) \, c_{63|54}(F_{6|54}, F_{3|54}) \, c_{62|543}(F_{6|543}, F_{2|543}) \, c_{75}(F_7, F_5) \\ & \cdot c_{76|5}(F_{7|5}, F_{6|5}) \, c_{73|56}(F_{7|56}, F_{3|56}). \end{aligned}$$

We will later derive a pair-copula decomposition for $F_{3|56}$. In preparation, we now use Algorithm 1 to derive pair-copula decompositions for f_{356} and f_{56}.

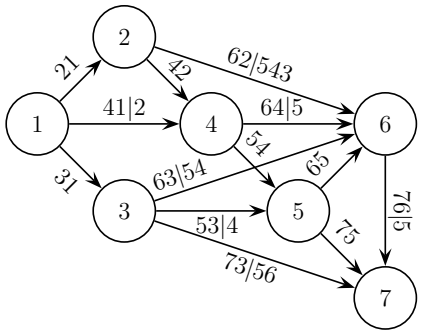
As a result of applying Algorithm 1 to f_{356} and f_{56}, we obtain

$$\begin{aligned} f_{356} &= \int_{\mathbb{R}^3} f_{6|543} \, f_{5|43} \, f_{4|21} \, f_{3|1} \, f_{2|1} \, f_1 \, \mathrm{d}\boldsymbol{x}_{124} \\ &= \int_{\mathbb{R}^3} f_6 \, c_{63|54}(F_{6|54}, F_{3|54}) \, c_{64|5}(F_{6|5}, F_{4|5}) \, c_{65}(F_6, F_5) \, f_5 \, c_{53|4}(F_{5|4}, F_{3|4}) \, c_{54}(F_5, F_4) \\ &\quad \cdot f_4 \, c_{41|2}(F_{4|2}, F_{1|2}) \, c_{42}(F_4, F_2) \, f_3 \, c_{31}(F_3, F_1) \, f_2 \, c_{21}(F_2, F_1) \, f_1 \, \mathrm{d}\boldsymbol{x}_{124} \end{aligned} \tag{3.6}$$

and $f_{56} = f_{6|5} \, f_5 = f_6 \, c_{65}(F_6, F_5) \, f_5$, respectively, see Table 3.1.

When can $\int_{-\infty}^{y} f_{\{v\}\cup K}(\boldsymbol{x}_{\{v\}\cup K}) \, \mathrm{d}x_v$ in Equation (3.3) be further simplified?

Let us now return to the conditional cdf in Equation (3.3). Setting $I \coloneqq \{v\} \cup K$, the numerator on the right hand side of Equation (3.3) takes the form $\int_{-\infty}^{y} f_I(\boldsymbol{x}_I) \, \mathrm{d}x_v$. Decompose $f_I(\boldsymbol{x}_I)$ according to Theorem 3.4, and let J denote the set of vertices corresponding to the newly added integration variables. Clearly, $J \subseteq \mathrm{An}(I) \setminus I$. If the old integration variable x_v does not appear as a conditioning variable in one of the pair copulas $C_{v,w|\mathrm{pa}(v;w)}$, $v \in I \cup J$, $w \in W_v$, in the decomposition of f_I, it may be possible to solve the integral with respect to x_v analytically. More precisely, let $J' \subseteq J$ and let $W' \subseteq W \coloneqq I_{-v} \cup J'$ be non-empty. Let $k \coloneqq |W'|$ and write $W' = \{w_1, \ldots, w_k\}$. Moreover, set $W'_{-i} \coloneqq \{w_1, \ldots, w_{i-1}\}$ for all $i \in \{1, \ldots k\}$. Assume that the (conditional) pair copula $C_{v,w_k|W'_{-k}}$ is available in the pair-copula decomposition of f, that is $W'_{-k} = \mathrm{pa}(v; w_k)$ or $W'_{-k} = \mathrm{pa}(w_k; v)$, and that (after



$$
\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\left[\begin{array}{ccccccc}
0 & 1 & 1 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 4 & 0 \\
0 & 0 & 0 & 0 & 2 & 3 & 3 \\
0 & 0 & 0 & 0 & 1 & 2 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right]
\end{array}
$$

Figure 3.2: A well-ordered (vertex labels) DAG $\mathcal{G}$ (left) with parent orderings $2 <_4 1$, $4 <_5 3$, $5 <_6 4 <_6 3 <_6 2$, $5 <_7 6 <_7 3$ specifying the pair copulas C_{21}, C_{31}, C_{42}, $C_{41|2}$, C_{54}, $C_{53|4}$, C_{65}, $C_{64|5}$, $C_{63|54}$, $C_{62|542}$, C_{75}, $C_{76|5}$, $C_{73|56}$ (edge labels), and corresponding representation matrix $A_{\mathcal{G}}$ (right).

f_{356}	I	v	S	$I_{-v} \subseteq \mathrm{pa}(v)$?	w	W	J
	$\{3,5,6\}$	6	$\emptyset$	✓	3	$\{3,4,5\}$	$\{4\}$
	$\{3,4,5\}$	5	$\emptyset$	✓	3	$\{3,4\}$	$\{4\}$
	$\{3,4\}$	4	$\emptyset$	✗	1	$\{1,2\}$	$\{1,2,4\}$
	$\{1,2,3\}$	3	$\emptyset$	✗	1	$\{1\}$	$\{1,2,4\}$
	$\{1,2\}$	2	$\emptyset$	✓	1	$\{1\}$	$\{1,2,4\}$
	$\{1\}$	1	$\emptyset$	–	–	$\emptyset$	$\{1,2,4\}$
f_{56}	I	v	S	$I_{-v} \subseteq \mathrm{pa}(v)$?	w	W	J
	$\{5,6\}$	6	$\emptyset$	✓	5	$\{5\}$	$\emptyset$
	$\{5\}$	5	$\{3,4\}$	–	–	$\emptyset$	$\emptyset$

Table 3.1: Vertices and vertex sets obtained during the application of Algorithm 1 to the pdfs f_{356} and f_{56} corresponding to the DAG $\mathcal{D}$ in Figure 3.2.

possible algebraic manipulation) f_I takes the form

$$\int_{\mathbb{R}^{|J'|}} f_v(x_v) \prod_{i=1}^{k} c_{v,w_i|W'_{-i}}\big(F_{v|W'_{-i}}(x_v \,|\, \boldsymbol{x}_{W'_{-i}}), F_{w_i|W'_{-i}}(x_{w_i} \,|\, \boldsymbol{x}_{W'_{-i}}) \,\big|\, \boldsymbol{x}_{W'_{-i}}\big)\, f_W(\boldsymbol{x}_W)\,\mathrm{d}\boldsymbol{x}_{J'}. \tag{3.7}$$

Then Fubini's theorem and Equation (2.5) yield that $\int_{-\infty}^{y} f_I(\boldsymbol{x}_I)\,\mathrm{d}x_v$ takes the form

$$\int_{\mathbb{R}^{|J'|}} h_{\underline{v},w_k|W'_{-k}}\big(F_{v|W'_{-k}}(y \,|\, \boldsymbol{x}_{W'_{-k}}), F_{w_k|W'_{-k}}(x_{w_k} \,|\, \boldsymbol{x}_{W'_{-k}}) \,\big|\, \boldsymbol{x}_{W'_{-k}}\big)\, f_W(\boldsymbol{x}_W)\,\mathrm{d}\boldsymbol{x}_{J'}, \tag{3.8}$$

where the integral with respect to x_v was replaced by an h-function which, by assumption, is available in the pair-copula decomposition of f. Note that some of the copula pdfs $c_{v,w_i|W'_{-i}}$, $i \in \{1,\dots,k-1\}$, in Equation (3.7) may not correspond to an edge in $\mathcal{D}$, but may instead be given implicitly by an integral over further variables, or may be equal to 1 due to a related Markov property of P, see also the example below. We need to take these special cases into account when checking the applicability of the inverse chain rule algorithmically.

It may sometimes also be useful to substitute $u_w \coloneqq F_w(x_w)$, $\mathrm{d}u_w = f_w(x_w)\,\mathrm{d}x_w$, for all $w \in J$ in the pair-copula decomposition of f_I, and thus to write

$$f_I(\boldsymbol{x}_I) = \int_{[0,1]^{|J|}} c_{I\cup J}\Big(\big(F_w(x_w)\big)_{w\in I}, \boldsymbol{u}_J\Big) \prod_{w\in I} f_w(x_w)\,\mathrm{d}\boldsymbol{u}_J.$$

A similar transformation can be applied to the denominator in Equation (3.3) if integration variables are present.

Example 3.3. Consider the integral $\int_{-\infty}^{\cdot} f_{356}\,\mathrm{d}x_3$ associated to the DAG $\mathcal{G}$ in Figure 3.2, which will later appear when deriving a pair-copula decomposition for $F_{3|56}$. Observing that

$$\begin{aligned}
&\int_{\mathbb{R}^2} f_4\, c_{41|2}(F_{4|2}, F_{1|2})\, c_{42}(F_4, F_2)\, f_3\, c_{31}(F_3, F_1)\, f_2\, c_{21}(F_2, F_1)\, f_1\,\mathrm{d}\boldsymbol{x}_{12} \\
&= \int_{\mathbb{R}^2} f_{1234}\,\mathrm{d}\boldsymbol{x}_{12} = f_{34} = f_4\, f_3\, c_{43}(F_4, F_3),
\end{aligned}$$

Equation (3.6) yields

$$\int_{-\infty}^{\cdot} f_{356}\,\mathrm{d}x_3 = \int_{-\infty}^{\cdot}\int_{\mathbb{R}} f_3\, c_{63|54}(F_{6|54}, F_{3|54})\, c_{53|4}(F_{5|4}, F_{3|4})\, c_{43}(F_4, F_3) \cdot f_6\, c_{64|5}(F_{6|5}, F_{4|5})\, c_{65}(F_6, F_5)\, f_5\, c_{54}(F_5, F_4)\, f_4\,\mathrm{d}x_4\,\mathrm{d}x_3.$$

Note that c_{43} is not available in the pair-copula decomposition of f. Since by (2.5)

$$\int_{-\infty}^{\cdot} f_3\, c_{63|54}(F_{6|54}, F_{3|54})\, c_{53|4}(F_{5|4}, F_{3|4})\, c_{43}(F_4, F_3)\,\mathrm{d}x_3 = h_{6\underline{3}|54}(F_{6|54}, F_{3|54}),$$

we can, however, simplify the integral with respect to x_3, and c_{43} vanishes. We get

$$\int_{-\infty}^{\cdot} f_{356}\,\mathrm{d}x_3 = \int_{\mathbb{R}} f_6\, h_{6\underline{3}|54}(F_{6|54}, F_{3|54})\, c_{64|5}(F_{6|5}, F_{4|5})\, c_{65}(F_6, F_5)\, f_5\, c_{54}(F_5, F_4)\, f_4\,\mathrm{d}x_4. \tag{3.9}$$

3.2.2 Pair-copula decompositions for conditional cdfs

Summing up, a pair-copula decomposition for the conditional cdf $F_{v|K}(\,\cdot\,|\,\boldsymbol{x}_K)$ in Equation (3.3) is obtained in three steps. First, we apply Theorem 3.4 to $f_{\{v\}\cup K}$ and f_K. Second, we possibly apply the inverse chain rule to the integral with respect to x_v in the numerator. Last, we cancel common factors like $\prod_{w\in K} f_w(x_w)$ in the numerator and the denominator. The procedure is summarised in Algorithm 2.

As is seen from Theorem 3.4 and Equation (3.8), the factorisation for $F_{v|K}(\,\cdot\,|\,\boldsymbol{x}_K)$ obtained from Algorithm 2 may contain some new conditional cdfs. This problem can, however, be solved inductively. Let w denote the maximal vertex in $\{v\}\cup K$ by the well-ordering of $\mathcal{D}$. Since Algorithm 2 only adds ancestors of $\{v\}\cup K$ as integration variables, all vertices involved in the new conditional cdfs are smaller than or equal to w by the well-ordering of $\mathcal{D}$. In particular, those conditional cdfs involving w are of the special form $F_{w|\mathrm{pa}(w;u)}(\,\cdot\,|\,\boldsymbol{x}_{\mathrm{pa}(w;u)})$ for some $u\in\mathrm{pa}(w)$, and can by Equation (2.4) iteratively be expressed as

$$F_{v|\mathrm{pa}(w;u)}(x_v|\boldsymbol{x}_{\mathrm{pa}(w;u)}) = h_{\underline{v},u|\mathrm{pa}(v;u)}\big(F_{v|\mathrm{pa}(v;u)}(x_v|\boldsymbol{x}_{\mathrm{pa}(v;u)}), F_{u|\mathrm{pa}(v;u)}(x_u|\boldsymbol{x}_{\mathrm{pa}(v;u)})\big|\boldsymbol{x}_{\mathrm{pa}(v;u)}\big).$$

Algorithm 2 Pair-copula decomposition of a conditional cdf.

Input Well-ordered DAG $\mathcal{D}$; set of parent orderings $\mathcal{O}$; vertex $v \in V$ (conditioned variable), vertex set $K \subseteq V_{-v}$ (conditioning variables).

Output Factorisation F. % conditional cdf $F_{v|K}(y \,|\, \boldsymbol{x}_K)$

% Exploit global $\mathcal{D}$-Markov property:
while $\exists\, w \in K \colon \{v\} \perp \{w\} \mid K_{-w}\, [(\mathcal{D}_{\mathrm{An}(\{v\} \cup K)})^m]$ **do**
    $K \leftarrow K_{-w}$;
end while
% Numerator:
$n \leftarrow \mathrm{Algorithm\_1}\big(\mathcal{D}, \mathcal{O}, \{v\} \cup K\big)$;
$n \leftarrow \int_{-\infty}^{y} n \,\mathrm{d}x_v$;
simplify n with inverse chain rule for variable x_v if possible;
% Denominator:
$d \leftarrow \mathrm{Algorithm\_1}\big(\mathcal{D}, \mathcal{O}, K\big)$;
% Conditional cdf $F_{v|K}(y \,|\, \boldsymbol{x}_K)$:
cancel common factors in n and d;
$F \leftarrow \frac{n}{d}$;

Hence, all vertices involved in the algorithmically more demanding new conditional cdfs are strictly smaller than w by the well-ordering of $\mathcal{D}$. Corresponding pair-copula decompositions for the new conditional cdfs can thus be computed inductively by again applying Algorithm 2. Since V is finite, the whole procedure terminates after finitely many steps, and the desired decomposition in terms of only univariate marginals and (conditional) pair copulas is obtained.

Overall, we observed that the problems of deriving pair-copula decompositions for a conditional cdf and a marginal pdf are deeply intertwined and can be solved by alternating iteration. Note that it is sufficient for our purposes to exploit only those conditional independence properties of P which follow directly from graph separation in $\mathcal{D}$ via the global $\mathcal{D}$-Markov property. Once a complete decomposition for f is obtained, the evaluation at $\boldsymbol{x} \in \mathbb{R}^d$ can be performed vertex-by-vertex and parent-by-parent along the well-ordering of $\mathcal{D}$. That is, given $v^* \in V$ and $w^* \in \mathrm{pa}(v^*)$, we first evaluate all terms corresponding to the marginals F_v and the pair copulas $C_{v,w|\mathrm{pa}(v;w)}$ for v smaller than v^* by the well-ordering of $\mathcal{D}$ and $w <_v w^*$ if $w^* \in \mathrm{pa}(v)$, before evaluating the terms corresponding to F_{v^*} and $C_{v^*,w^*|\mathrm{pa}(v^*;w^*)}$.

Example 3.4. For the DAG $\mathcal{G}$ in Figure 3.2, we sketch how to apply Algorithm 2 to obtain a pair-copula decomposition for $F_{3|56}$. Note that $\mathcal{G}$ contains the edges $3 \to 5$ and $3 \to 6$, which is why neither 5 nor 6 can be removed from the conditioning set. We get $F_{3|56} = \int_{-\infty}^{\cdot} \frac{f_{356}}{f_{56}} \, \mathrm{d}x_3$. Applying our previous results for f_{356} and f_{56}, respectively, we further have

$$F_{3|56} = \frac{\int_{\mathbb{R}} f_6 \, h_{63|54}(F_{6|54}, F_{3|54}) \, c_{64|5}(F_{6|5}, F_{4|5}) \, c_{65}(F_6, F_5) \, f_5 \, c_{54}(F_5, F_4) \, f_4 \, \mathrm{d}x_4}{f_6 \, c_{65}(F_6, F_5) \, f_5},$$

see Equation (3.9). Thus, by cancelling common factors, we finally obtain

$$F_{3|56} = \int_{\mathbb{R}} h_{63|54}(F_{6|54}, F_{3|54}) \, c_{64|5}(F_{6|5}, F_{4|5}) \, c_{54}(F_5, F_4) \, f_4 \, \mathrm{d}x_4.$$

The complete pair-copula decomposition for f is given in Appendix B.

3.3 Simulation and ML estimation in pair-copula Bayesian networks

Given (conditional) pair copulas $C_{v,w|\mathrm{pa}(v;w)}(\,\cdot\,,\cdot\,;\boldsymbol{\theta}_{v,w|\mathrm{pa}(v;w)})$, $v \in V$, $w \in \mathrm{pa}(v)$, with joint parameter vector $\boldsymbol{\theta} \coloneqq (\boldsymbol{\theta}_{v,w|\mathrm{pa}(v;w)})_{v \in V, w \in \mathrm{pa}(v)} \in \boldsymbol{\Theta}$, above construction yields a d-variate copula model, which we will denote by $\{C_{\mathcal{D},\mathcal{O},\boldsymbol{\theta}} \,|\, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$. Note that for computational convenience, we again make the *simplifying assumption* of constant conditional copulas. Together with families of univariate marginals, $\{C_{\mathcal{D},\mathcal{O},\boldsymbol{\theta}} \,|\, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ constitutes a statistical model which merges the advantages of graphical Markov modelling with the distributional flexibility of the pair-copula approach. We will refer to such a model as a *pair-copula Bayesian network* (PCBN).

3.3.1 Simulation

Write $V = \{v_1, \ldots, v_d\}$ according to the well-ordering of $\mathcal{D}$, and for $i \in \{1, \ldots, d\}$ set $V_{-i} \coloneqq \{v_1, \ldots, v_{i-1}\}$. A sample $\boldsymbol{u} = (u_{v_1}, \ldots, u_{v_d}) \in [0,1]^d$ from a fully specified PCBN with uniform univariate margins is obtained by simulating d independent

uniform $[0,1]$ variables $x_1, \dots, x_d$ and applying the quantile transformations

$$\begin{aligned} u_{v_1} &\coloneqq x_1, \\ u_{v_2} &\coloneqq F^{-1}_{v_2|v_1}(x_2 \,|\, u_{v_1}; \boldsymbol{\theta}), \\ u_{v_3} &\coloneqq F^{-1}_{v_3|V_{-3}}(x_3 \,|\, \boldsymbol{u}_{V_{-3}}; \boldsymbol{\theta}), \\ &\vdots \\ u_{v_d} &\coloneqq F^{-1}_{v_d|V_{-d}}(x_d \,|\, \boldsymbol{u}_{V_{-d}}; \boldsymbol{\theta}). \end{aligned}$$

The order in which the components of $\boldsymbol{u}$ are generated is given by the well-ordering of $\mathcal{D}$. Solving transformation equation i for x_i, we have by the local $\mathcal{D}$-Markov property

$$x_i = F_{v_i|V_{-i}}(u_{v_i} \,|\, \boldsymbol{u}_{V_{-i}}; \boldsymbol{\theta}) = F_{v_i|\mathrm{pa}(v_i)}(u_{v_i} \,|\, \boldsymbol{u}_{\mathrm{pa}(v_i)}; \boldsymbol{\theta}), \quad i \in \{1, \dots, d\}. \tag{3.10}$$

Now assume that $\mathrm{pa}(v_i) \neq \emptyset$, and let w denote the largest vertex in $\mathrm{pa}(v_i)$ by the parent ordering $<_{v_i}$. Then Equations (3.10) and (2.4) yield

$$x_i = h_{\underline{v_i}, w|\mathrm{pa}(v_i;w)}\big(F_{v_i|\mathrm{pa}(v_i;w)}(u_{v_i} \,|\, \boldsymbol{u}_{\mathrm{pa}(v_i;w)}; \boldsymbol{\theta}), F_{w|\mathrm{pa}(v_i;w)}(u_w \,|\, \boldsymbol{u}_{\mathrm{pa}(v_i;w)}; \boldsymbol{\theta}); \boldsymbol{\theta}\big). \tag{3.11}$$

Since u_{v_i} is only contained in the first argument $F_{v_i|\mathrm{pa}(v_i;w)}(...)$ on the right hand side of Equation (3.11), we obtain by induction that the only inverse functions needed in the computation of u_{v_i} are the inverse h-functions $h^{-1}_{\underline{v_i}, w^*|\mathrm{pa}(v_i, w^*)}$, $w^* \in \mathrm{pa}(v_i)$.

Example 3.5. Consider again the well-ordered DAG $\mathcal{D}$ in Figure 1.2 with parent ordering $2 <_4 3$. Moreover, let $x_1, \dots, x_4 \in [0,1]$. Straightforward application of the results in Example 3.1 yields the sampling equations

$$\begin{aligned} u_1 &\coloneqq x_1, \\ u_2 &\coloneqq h^{-1}_{\underline{2}1}(x_2, u_1; \boldsymbol{\theta}_{21}), \\ u_3 &\coloneqq h^{-1}_{\underline{3}1}(x_3, u_1; \boldsymbol{\theta}_{31}), \\ u_4 &\coloneqq h^{-1}_{\underline{4}2}\left(h^{-1}_{\underline{4}3|2}\left(x_4, \int_0^1 c_{21}(u_2, y_1; \boldsymbol{\theta}_{21})\, h_{\underline{3}1}(u_3, y_1; \boldsymbol{\theta}_{31})\, \mathrm{d}y_1; \boldsymbol{\theta}_{43|2}\right), u_2; \boldsymbol{\theta}_{42}\right). \end{aligned}$$

3.3.2 ML estimation

Let $\boldsymbol{u} = (\boldsymbol{u}^1, \dots, \boldsymbol{u}^n)$, $n \in \mathbb{N}$, be a realisation of a sample of i.i.d. observations $\boldsymbol{U}^1, \dots, \boldsymbol{U}^n$ from a random variable $\boldsymbol{U}$ on $[0,1]^d$ with copula family $\{C_{\mathcal{D},\mathcal{O},\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ and uniform univariate margins. The restriction to uniform margins is made along the same lines as for vine copula models. By Equation (3.2), the log-likelihood function $l(\boldsymbol{\theta}; \boldsymbol{u})$ takes the form

$$\sum_{k=1}^{n} \sum_{v \in V} \sum_{w \in \mathrm{pa}(v)} \log c_{v,w|\mathrm{pa}(v;w)} \Big(F_{v|\mathrm{pa}(v;w)}\big(u_v^k \mid \boldsymbol{u}_{\mathrm{pa}(v;w)}^k; \boldsymbol{\theta}\big), F_{w|\mathrm{pa}(v;w)}\big(u_w^k \mid \boldsymbol{u}_{\mathrm{pa}(v;w)}^k; \boldsymbol{\theta}\big); \boldsymbol{\theta} \Big). \tag{3.12}$$

ML estimation of the parameters in Equation (3.12) can be performed using a stepwise approach similar to the one discussed in Section 2.3. The only difference to vine copula models is that we iterate over the vertices of $\mathcal{D}$ and their respective parents instead of over the trees of an R-vine. Hence again, in a first step, *sequential ML estimates* are computed and in a second step, using the sequential ML estimates as starting values, *joint ML estimates* $\widehat{\boldsymbol{\theta}}_{v,w|\mathrm{pa}(v;w)}$, $v \in V$, $w \in \mathrm{pa}(v)$, are inferred.

Example 3.6. For the PCBN in Example 3.5 we obtain the log-likelihood function

$$\begin{aligned} l(\boldsymbol{\theta}; \boldsymbol{u}) = \sum_{k=1}^{n} & \log c_{21}\big(u_2^k, u_1^k; \boldsymbol{\theta}_{21}\big) + \log c_{31}\big(u_3^k, u_1^k; \boldsymbol{\theta}_{31}\big) + \log c_{42}\big(u_4^k, u_2^k; \boldsymbol{\theta}_{42}\big) \\ & + \log c_{43|2}\Big(h_{\underline{4}2}\big(u_4^k, u_2^k; \boldsymbol{\theta}_{42}\big), \int_0^1 c_{21}\big(u_2^k, y_1; \boldsymbol{\theta}_{21}\big)\, h_{\underline{3}1}\big(u_3^k, y_1; \boldsymbol{\theta}_{31}\big)\, \mathrm{d}y_1; \boldsymbol{\theta}_{43|2} \Big). \end{aligned}$$

3.4 Likelihood inference: A simulation study

An appealing feature of PCBNs is their flexibility gained from using bivariate copulas as building blocks only. In particular, these models can accommodate distributions other than the multivariate normal, which is a desirable property in many statistical applications, see, for instance, McNeil et al. (2005, Section 3.1.4). We now investigate these models' aptitude for likelihood inference. Again, we restrict our considerations to PCBNs with uniform $[0,1]$ univariate margins.

Copula		1	2	3	4	5	6
C_{21}	C	0.25	0.75	0.75	0.25	0.25	0.25
C_{31}	G	0.25	0.75	0.25	0.75	0.25	0.25
C_{42}	t	0.25	0.75	0.25	0.25	0.75	0.25
$C_{43\|2}$	N	0.25	0.75	0.25	0.25	0.25	0.75
Copula		7	8	9	10	11	12
C_{21}	C	0.25	0.75	0.75	0.25	0.25	0.25
C_{31}	C	0.25	0.75	0.25	0.75	0.25	0.25
C_{42}	C	0.25	0.75	0.25	0.25	0.75	0.25
$C_{43\|2}$	C	0.25	0.75	0.25	0.25	0.25	0.75

Table 3.2: Selected pair-copula families and values of Kendall's τ for C_{21}, C_{31}, C_{42}, and $C_{43|2}$. Copulas were chosen from the Clayton (C), Gumbel (G), Gaussian (N), and Student's t (t) pair-copula families.

3.4.1 Simulation setup

We examined the small sample performance of PCBNs by a simulation study. To this end, we drew samples from two PCBNs with conditional independence properties given by the DAG $\mathcal{D}$ in Figure 1.2. More precisely, these PCBNs emerge from two different choices of pair-copula families involved. In either setting, we considered 6 different parameter configurations, resulting in the 12 simulation scenarios described in Table 3.2. The copula families used (Clayton, Gumbel, Gaussian, and Student's t) exhibit notable differences in their dependence structures and tail behaviours. Within each copula family, a range of values of Kendall's τ as a dependence measure can be obtained by suitable parameter choices. We considered two scenarios in which all selected copulas have $\tau = 0.25$, and two scenarios with $\tau = 0.75$. In each of the 8 remaining scenarios, one copula has $\tau = 0.75$, while all other copulas have $\tau = 0.25$. The degrees-of-freedom parameters of the selected Student's t copulas were set to $\nu = 5$ in order to allow for heavy-tailed dependence. See Table 3.3 for the relations between copula parameters, TDCs, and Kendall's τ.

In each of the 12 scenarios we performed $N = 100$ simulation runs, and in each simulation run we generated $n = 1{,}000$ i.i.d. observations. The sampling procedure

Copula	Clayton		Gumbel		Gauss		Student	
Parameter(s)	0.67	6.00	1.33	4.00	0.38	0.92	0.38, 5	0.92, 5
Kendall's τ	0.25	0.75	0.25	0.75	0.25	0.75	0.25	0.75
Lower TDC λ_{L}	0.35	0.89	0.00	0.00	0.00	0.00	0.15	0.15
Upper TDC λ_{U}	0.00	0.00	0.32	0.81	0.00	0.00	0.64	0.64

Table 3.3: Parameters, Kendall's correlation coefficients, and TDCs of the pair copulas used in the simulation study. See also Table 2.1.

used was outlined in Example 3.5. For each of the 1,200 runs, we numerically calculated joint ML estimates of the parameters of the underlying PCBN (as specified in Table 3.2) and compared them to the respective true parameters. In order to summarise ML estimates within each scenario, we used a common scale by converting all parameter estimates to estimates of Kendall's τ, as described in Table 2.1. An overview of our results in terms of empirical bias and mean squared error (MSE) with respect to the true value of Kendall's τ is given in Figure 3.5.

PCBNs may be viewed as generalisations of Gaussian Bayesian networks as presented in Section 1.3. By choosing all univariate margins as well as all pair-copula families in a PCBN based on a DAG $\mathcal{D}$ to be Gaussian, we obtain the Gaussian Bayesian network based on $\mathcal{D}$. The Gaussian DAG PCC corresponding to the DAG of our simulation study, for instance, is a four-variate Gaussian copula with correlation matrix

$$\boldsymbol{R} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & R_{14} \\ \rho_{12} & 1 & \rho_{12}\,\rho_{13} & \rho_{24} \\ \rho_{13} & \rho_{12}\,\rho_{13} & 1 & R_{34} \\ R_{14} & \rho_{24} & R_{34} & 1 \end{bmatrix},$$

where ρ_{12}, ρ_{13}, and ρ_{24} are the correlations implied by the pair copulas C_{21}, C_{31}, and C_{42}. Using the conditional correlation $\rho_{34|2}$ implied by the copula $C_{43|2}$, we may represent the correlations R_{14} and R_{34} as

$$R_{14} = \rho_{12}\,\rho_{24} + \frac{\rho_{13}\,\rho_{34|2}\,(1-\rho_{12}^2)\sqrt{1-\rho_{24}^2}}{\sqrt{1-\rho_{12}^2\,\rho_{13}^2}},$$

$$R_{34} = \rho_{12}\,\rho_{13}\,\rho_{24} + \rho_{34|2}\sqrt{1-\rho_{12}^2\,\rho_{13}^2}\sqrt{1-\rho_{24}^2}.$$

This representation is based on the conditional independence properties specified by the DAG $\mathcal{D}$ in Figure 1.2 and the iterative formula for partial correlations given in Equation (1.4). The latter is applicable since partial and conditional correlations coincide for normal distributions (Whittaker, 1990, Section 6.2). Fitting this *Gaussian PCBN* to our simulated data sets allows us to compare the estimated pair-copula parameters to the true parameters of the generating models. To ensure comparability, we again transform estimated parameters to estimates of Kendall's τ. Error estimates in the Gaussian PCBN are interpreted as in the true model, which we will henceforth refer to as the *non-Gaussian PCBN*. An overview of these error estimates is given in Figure 3.5.

As stated in Section 3.1, the PCBN corresponding to the DAG $\mathcal{D}$ of our simulation study cannot be represented by an R-vine. A D-vine featuring the same unconditional pair copulas C_{12}, C_{13}, and C_{24} as our DAG is given in Figure 3.3. It specifies a D-vine copula model that approximates our PCBN. In the second and third tree, this D-vine comprises the conditional pair copulas $C_{23|1}$, $C_{14|2}$, and $C_{34|12}$. In order to study how well this D-vine copula model approximates the given PCBN, we performed the following procedure in each of the 1,200 runs of our simulation study. First, we selected pair-copula families for C_{12}, C_{13}, C_{24}, $C_{23|1}$, $C_{14|2}$, and $C_{34|12}$, choosing from the four copula families described in Table 3.3 and the independence copula. More precisely, we first computed sequential ML estimates of the parameters $\boldsymbol{\theta}_{12}$, $\boldsymbol{\theta}_{13}$, $\boldsymbol{\theta}_{24}$, $\boldsymbol{\theta}_{14|2}$, $\boldsymbol{\theta}_{23|1}$, and $\boldsymbol{\theta}_{34|12}$ for the Clayton, Gumbel, Gaussian, and Student's t copula, respectively, and then used the AIC to identify the most appropriate copula families. We included C_{12}, C_{13}, and C_{24}, of which the true families are known, in this procedure to be able to judge the reliability of our selection criterion. Table 3.4 gives an overview of how often each copula family was selected. In almost all simulation runs, the families of C_{12}, C_{13}, and C_{24} were identified correctly. If $C_{23|1}$ is the independence copula in this model, then the D-vine in Figure 3.3 yields a PCC satisfying $U_2 \perp\!\!\!\perp U_3 \mid U_1$ but, in general, fails to satisfy $U_1 \perp\!\!\!\perp U_4 \mid \boldsymbol{U}_{23}$ as specified by $\mathcal{D}$. If the independence copula does not appear in the

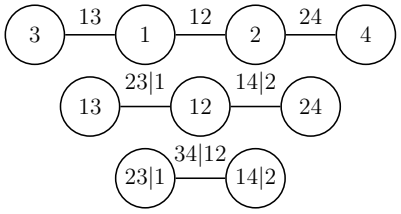


Figure 3.3: A four-variate D-vine having the unconditional pair copulas in common with the DAG PCC derived from the DAG $\mathcal{D}$ in Figure 1.2.

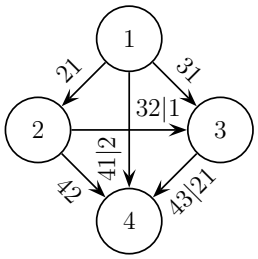


Figure 3.4: A complete DAG specifying the same PCC as the D-vine in Figure 3.3 given that none of the associated pair copulas is the independence copula. Parents were ordered according to $1 <_3 2$ and $2 <_4 1 <_4 3$.

D-vine copula model, we obtain the same log-likelihood as in the model specified by the complete DAG in Figure 3.4, which is hence an example of a PCBN that can be represented by a D-vine. Note, however, that interest with Bayesian networks lies in capturing conditional independence assumptions and therefore rather in DAGs with missing edges than in complete DAGs. The approximating D-vine copula model can be viewed as structurally misspecified model for the given data. Given a choice of pair-copula families, we then computed joint ML parameter estimates in the D-vine copula model and compared the maximised log-likelihoods to those obtained in the Gaussian and the non-Gaussian PCBN. The results are displayed in Figure 3.6.

3.4.2 Results

Since the Gaussian PCBN can be viewed as a misspecified model for the data given in scenarios 1 through 12, it is not surprising that error estimates in this model are generally higher than those in the non-Gaussian PCBNs. In fact, Figure 3.5

Copula		1	2	3	4	5	6	7	8	9	10	11	12
C_{12}	C	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	G, N, t, I: all 0												
C_{13}	C	0	0	0	0	0	0	**100**	**100**	**100**	**100**	**100**	**100**
	G	**100**	**100**	**99**	**100**	**99**	**98**	0	0	0	0	0	0
	N	0	0	1	0	1	0	0	0	0	0	0	0
	t	0	0	0	0	0	2	0	0	0	0	0	0
	I	0	0	0	0	0	0	0	0	0	0	0	0
C_{24}	C	0	0	0	0	0	0	**100**	**100**	**100**	**100**	**100**	**100**
	G	0	0	2	0	0	1	0	0	0	0	0	0
	t	**100**	**100**	**98**	**100**	**100**	**99**	0	0	0	0	0	0
	N, I: all 0												
$C_{14\|2}$	C	3	0	4	1	1	0	95	29	39	100	98	98
	G	36	86	36	4	36	98	0	0	8	0	0	0
	N	58	0	43	89	54	1	2	6	18	0	2	1
	t	3	14	4	6	8	1	3	65	4	0	0	1
	I	0	0	13	0	1	0	0	0	31	0	0	0
$C_{23\|1}$	C	3	3	1	6	8	2	4	2	5	7	6	3
	G	3	3	3	4	5	4	3	1	5	4	4	2
	N	13	9	7	8	10	10	16	7	10	7	11	14
	t	5	1	0	2	5	4	1	1	4	3	1	1
	I	**76**	**84**	**89**	**80**	**72**	**80**	**76**	**89**	**76**	**79**	**78**	**80**
$C_{34\|12}$	C	0	0	0	13	0	0	100	81	100	31	99	100
	G	0	0	0	3	0	0	0	0	0	2	0	0
	N	93	22	95	78	93	95	0	0	0	62	0	0
	t	7	78	5	6	7	5	0	19	0	5	1	0
	I	0	0	0	0	0	0	0	0	0	0	0	0

Table 3.4: For each scenario (column) and run we identified the copula families with optimal AIC. The resulting frequencies are given. Frequencies of the true copula families for C_{12}, C_{13}, C_{24}, $C_{23|1}$ appear in bold. Copulas were chosen from the Clayton (C), Gumbel (G), Gaussian (N), and Student's t (t) pair-copula families, and the independence copula (I). As an example, the top left entry of the table states that in all of the 100 runs of scenario 1, C as a choice for C_{12} yields a higher AIC than all other families considered (G, N, t, I). Also, C is the true copula family for C_{12} in this scenario.

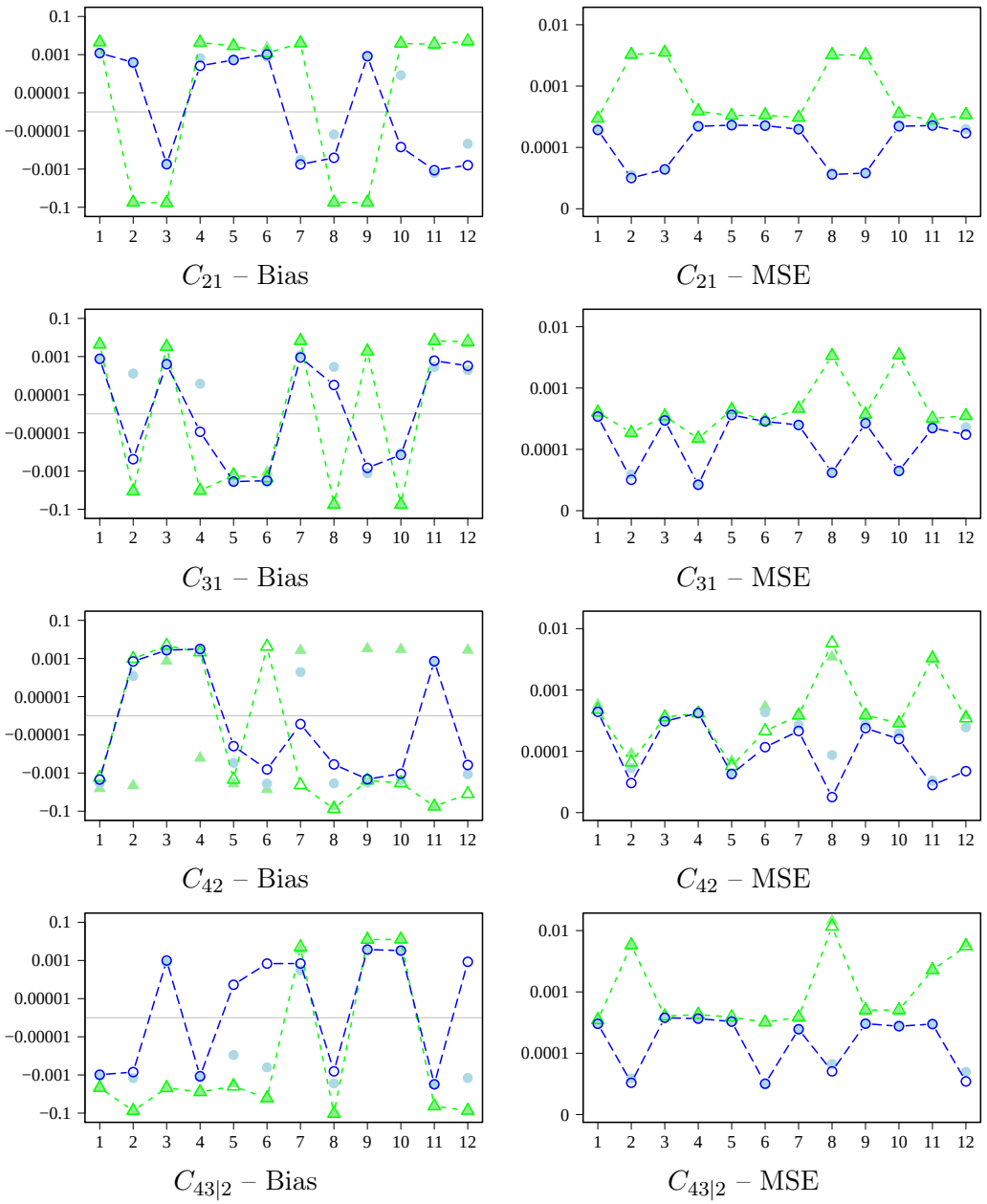


Figure 3.5: Bias (left) and MSE (right) of estimates of Kendall's τ associated with copulas C_{21}, C_{31}, C_{42}, $C_{43|2}$. The vertical axes use a transformed log-scale for better visibility. In each of the 12 scenarios (horizontal axis) described in Table 3.2, parameter estimates were obtained in the true (circle) and in the Gaussian (triangle) PCBN using sequential (solid blue or green) and joint (outline blue or green) ML estimation.

shows that the estimates of Kendall's τ obtained in the Gaussian PCBN are considerably worse than those obtained in the non-Gaussian model. Differences between these models with regard to bias and MSE for Kendall's τ are more pronounced in scenarios featuring high correlations and asymmetric tail dependence. The smallest difference in performance is thus found in the low-correlation scenarios 1 and 7, whereas in the high-correlation scenarios 2 and 8 the Gaussian PCBN fails by a huge amount. For instance, in scenario 8 the ratios of biases for Kendall's τ in the non-Gaussian and Gaussian PCBN, respectively, are of order 10^3 or higher. The corresponding ratios of MSEs are of order 10^2 or higher. Note that Figure 3.5, with its focus on estimation of Kendall's τ, presents only one aspect in the comparison between non-Gaussian and Gaussian PCBNs. For instance, estimation of the degrees of freedom of a Student's t copula is neglected in this figure. It is clear that Gaussian PCBNs are useless when interest lies in the estimation of TDCs, say.

Since DAG and D-vine copula models have different parameter sets, direct comparisons of parameter estimates are infeasible. We hence compared maximised log-likelihoods for the Gaussian and the non-Gaussian PCBN and the D-vine copula model. Figure 3.6 shows that maximised log-likelihoods in the non-Gaussian PCBNs are roughly 50% higher than those obtained in their Gaussian counterparts. Even though the Gaussian PCBN in scenarios 1 to 6 has one parameter less than its non-Gaussian competitor, this difference in log-likelihood clearly shows the latter model's superiority. Figure 3.6 also shows the performance of the approximating D-vine copula model. As this model and the non-Gaussian PCBN share the same unconditional pair copulas, it is not surprising that the maximised log-likelihoods in these models differ mainly in those scenarios with high values of $\boldsymbol{\theta}_{43|2}$, namely, scenarios 2, 6, 8, and 12. In other words, the effects of structural misspecification are primarily noticeable in those 4 scenarios. Note, however, that the D-vine copula model involves a higher number of parameters than the non-Gaussian PCBN and that its maximised log-likelihood is always slightly inferior to that of the latter model.

The average computation time for joint ML estimation in the non-Gaussian PCBN was 12 seconds in scenarios 1 through 6 and 5 seconds in scenarios 7 through 12 on

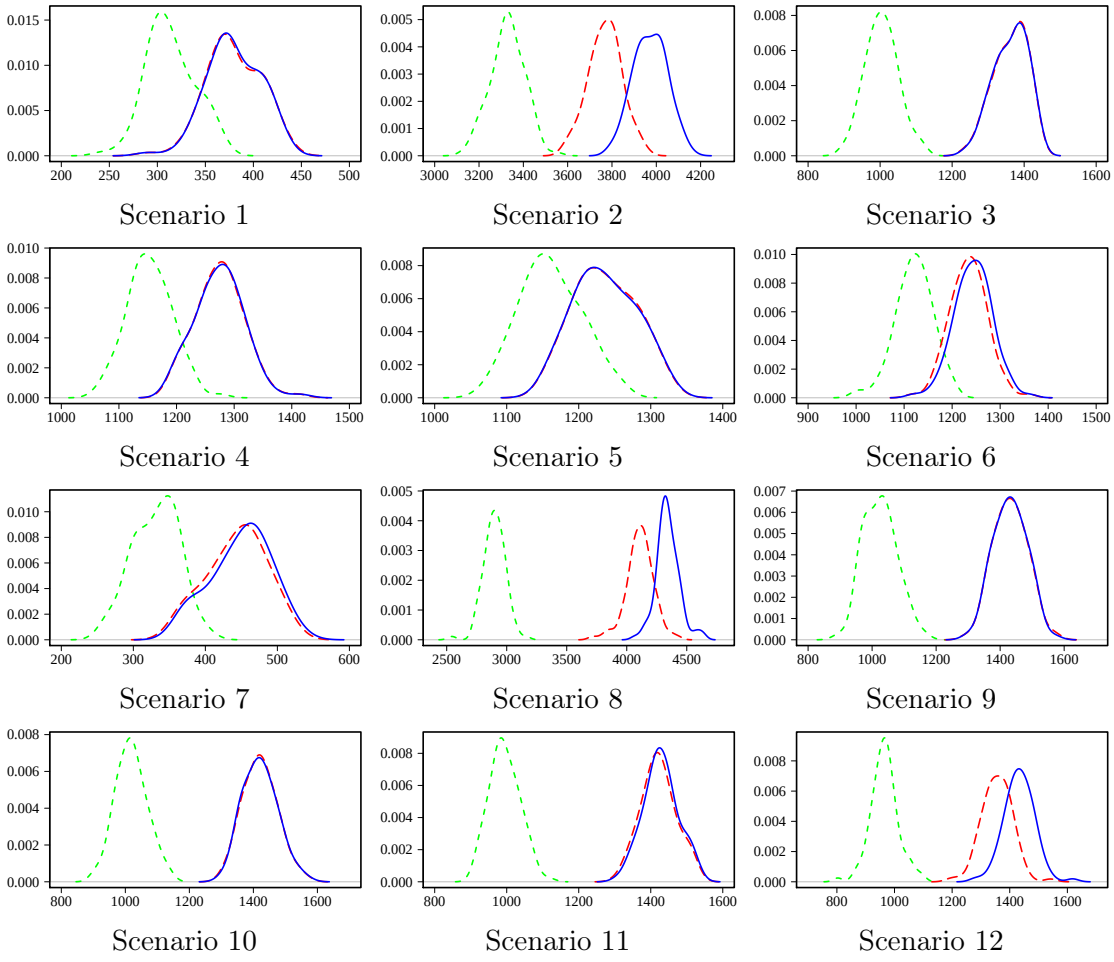


Figure 3.6: Kernel density estimates of the maximised log-likelihoods in 100 runs for each scenario (Table 3.2) based on the Gaussian PCBN (dashed), the non-Gaussian PCBN (solid), and the D-vine copula model (longdash).

a 2 GHz dual-core computer with 2 GB of RAM. Fitting the D-vine copula model instead reduced computation time by about 70% in scenarios 1, 3, 4, 5, and 6, and by about 90% in scenarios 7, 9, 10, 11, and 12. This difference in computation time is due to the numerical integration involved in computing joint ML estimates for the non-Gaussian PCBN. In the high-correlation scenarios 2 and 8, however, ML estimation in both models was performed equally fast. Overall, there is an increase in computation time when using the non-Gaussian PCBN instead of the D-vine copula model, but it is small compared to the associated gain in statistical precision.

Besides maximised log-likelihoods, we also investigated how well the three models capture rank correlations of the bivariate margins $\boldsymbol{U}_{14}$, $\boldsymbol{U}_{23}$, and $\boldsymbol{U}_{34}$, which were not directly included in our DAG and D-vine PCCs. To this end, we performed the following procedure in each run. First, we computed estimates $\widehat{\tau}_{14}$, $\widehat{\tau}_{23}$, and $\widehat{\tau}_{34}$ of Kendall's τ for the three margins. Then we generated a sample of $n = 1{,}000$ i.i.d. observations from the Gaussian PCBN, the non-Gaussian PCBN, and the D-vine copula model, respectively, using the joint ML parameter estimates obtained before. For each of these samples we again computed estimates of Kendall's τ and compared the results to $\widehat{\tau}_{14}$, $\widehat{\tau}_{23}$, and $\widehat{\tau}_{34}$, respectively. Figure 3.7 gives an overview of our findings in terms of empirical bias and MSE for each scenario. The patterns described by these error estimates resemble those of the maximised log-likelihoods summarised above, and hence yield similar conclusions. The range of bias and MSE values is roughly the same as in Figure 3.5.

In order to check whether our findings are valid for smaller sample sizes, we conducted an additional simulation study in which each simulation run contained $n = 500$ (as opposed to $n = 1{,}000$) observations. As far as estimation of Kendall's τ is concerned, we found biases similar to those in Figures 3.5 and 3.7, while MSEs tended to be twice as high as those given in Figures 3.5 and 3.7. The comparison between Gaussian and non-Gaussian PCBNs may be summarised along the same lines as above. By comparison with its competitors, the performance of the vine copula model decreases for smaller sample size and is subject to higher variability.

The main conclusion from our simulation study is that non-Gaussian PCBNs are

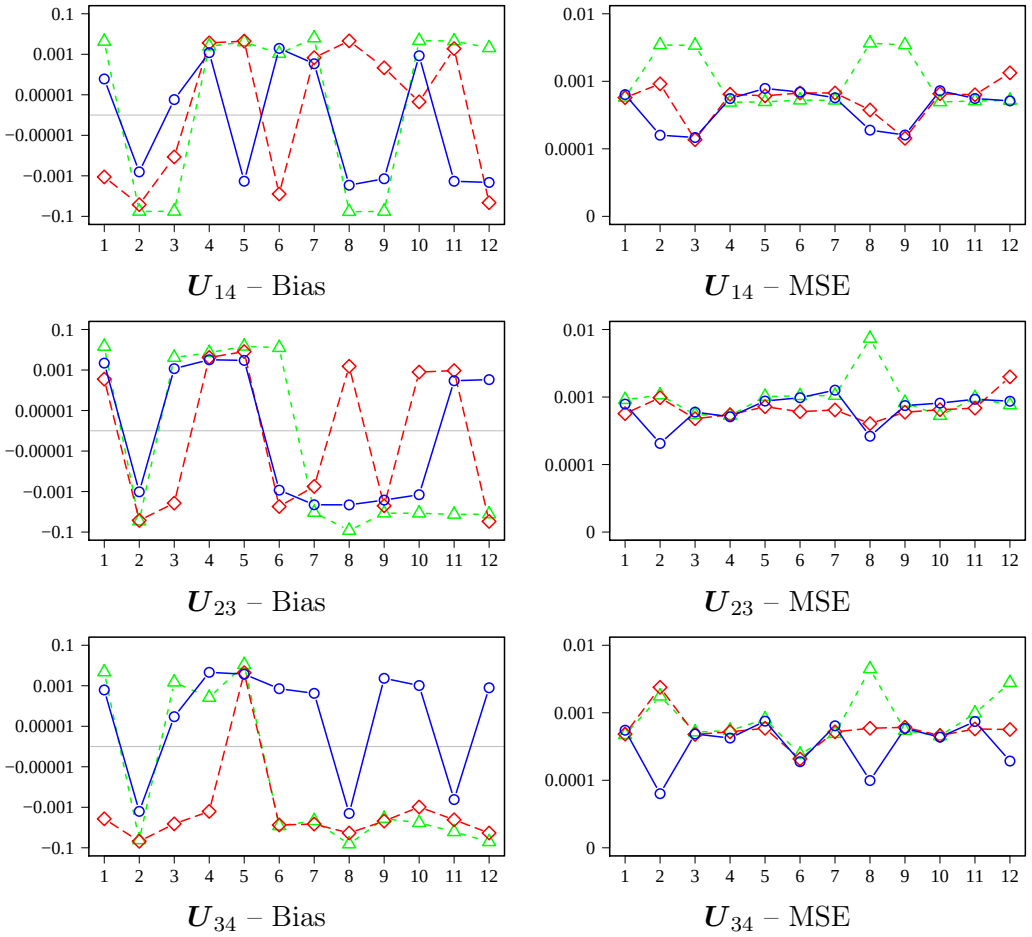


Figure 3.7: Bias (left) and MSE (right) of estimates of Kendall's τ associated with margins $\boldsymbol{U}_{14}$, $\boldsymbol{U}_{23}$, $\boldsymbol{U}_{34}$. The vertical axes use a transformed log-scale for better visibility. In each of the 12 scenarios (horizontal axis) (Table 3.2), estimates of Kendall's τ were compared to estimates of Kendall's τ obtained from samples generated from the Gaussian PCBN (triangle), the non-Gaussian PCBN (circle), and the D-vine copula model (diamond).

capable of capturing features in data that neither Gaussian PCBNs nor vine copula models can reflect. Gaussian PCBNs exhibit particularly poor performance in presence of non-normal tail behaviour. Vine copula models are not sufficiently flexible to observe certain Markov properties. Both non-normality and Markov properties are easily incorporated into our non-Gaussian PCBNs.

4 Model selection

Model selection for PCBNs involves estimation of the DAG $\mathcal{D} = (V, E)$, selection of the set $\mathcal{O}$ of parent orderings, and selection of the pair-copula families for $C_{v,w|\mathrm{pa}(v;w)}$, $v \in V$, $w \in \mathrm{pa}(v)$. In this chapter, we investigate structure estimation in PCBNs using the PC algorithm (Spirtes and Glymour, 1991). To this end, we introduce a novel test for conditional independence that is based on vine copula models. The performance of our structure estimation procedure is examined in a simulation study. Moreover, we develop routines for selecting $\mathcal{O}$ and pair-copula families.

4.1 Structure estimation

The first task of modelling the joint distribution of a given set of variables with a Bayesian network is to identify the DAG $\mathcal{D} = (V, E)$ specifying the Markov structure of the variables. A convenient approach to defining $\mathcal{D}$ is the use of *expert knowledge*. However, the scope of this approach is rather limited since expert knowledge is often incomplete or unavailable. Data-driven *structure estimation algorithms* provide a computer-based alternative to elicited expert knowledge. Robinson (1973) has shown that the number n_d of DAGs on $d := |V|$ labelled vertices is given by the recurrence equation

$$n_0 = 1, \qquad n_d = \sum_{k=1}^{d} (-1)^{k-1} \binom{d}{k} 2^{k(d-k)} \, n_{d-k}.$$

Since n_d grows super-exponentially in d, a systematic trial of all possible DAGs on V is infeasible, and thus efficient searching algorithms are required. A considerable number of structure estimation algorithms has been proposed over the last two decades, see Neapolitan (2003, Chapters 8 – 11) and Koller and Friedman (2009,

Chapter 18) for an overview. The majority of these algorithms follow one of the two estimation approaches predominant in the literature: the *constraint-based* and the *score-and-search-based* approach. In the constraint-based approach, $\mathcal{D}$ is inferred from a series of conditional independence tests. In the score-and-search-based approach, $\mathcal{D}$ is found by optimising a given scoring function—like AIC or BIC—over a suitable search space, for instance, the space of all DAGs or the space of all Markov-equivalence classes. Besides, there exist hybrid algorithms which combine both approaches. Unfortunately, available implementations of aforementioned algorithms are mainly confined to discrete or Gaussian models and are hence not suited for our non-Gaussian continuous Bayesian networks.

4.1.1 The PC algorithm

We will provide a structure estimation algorithm that is particularly suited to finding the DAG $\mathcal{D}$ underlying a non-Gaussian continuous Bayesian network. Our algorithm is a version of one of the most popular constraint-based estimation algorithms, the *PC algorithm* (named after its inventors *P*eter Spirtes and *C*lark Glymour), see Spirtes and Glymour (1991) and Spirtes et al. (2000, Section 5.4.2). To fix notation and for the reader's convenience, we now recall the PC algorithm. Let P be an absolutely continuous $\mathcal{D}$-Markovian probability measure on $[0,1]^d$ with uniform univariate margins. The restriction to uniform margins is made along the same lines as in Section 3.3. Moreover, let $\boldsymbol{u} = (\boldsymbol{u}^1, \dots, \boldsymbol{u}^n)$, $n \in \mathbb{N}$, be a realisation of a sample of i.i.d. observations $\boldsymbol{U}^1, \dots, \boldsymbol{U}^n$ from a random variable $\boldsymbol{U}$ distributed as P. The PC algorithm for estimating $\mathcal{D}$ from $\boldsymbol{u}$ involves three major steps in which the complete UG $\mathcal{G}$ on V is gradually transformed into a CG $\mathcal{G}^*$ on V, which is supposed to be the essential graph $\mathcal{D}^e$ corresponding to the Markov-equivalence class $[\mathcal{D}]$ of $\mathcal{D}$. The resulting CG $\mathcal{G}^*$ can then be extended to a DAG as outlined in Section 1.2.

In the first step of the PC algorithm, a series of tests for conditional independence is performed on $\boldsymbol{u}$. More precisely, for all distinct vertices $i, j \in V$ and chosen vertex sets $K \subseteq V \setminus \{i, j\}$, the null hypothesis $\mathrm{H}_0 \colon U_i \perp\!\!\!\perp U_j \mid \boldsymbol{U}_K$ is tested against the general alternative $\mathrm{H}_1 \colon U_i \not\perp\!\!\!\perp U_j \mid \boldsymbol{U}_K$ of conditional dependence. Given a suitable

independence test of choice, we denote the test decision at significance level $\alpha \in (0,1)$ by $T_\alpha(\boldsymbol{u}_i, \boldsymbol{u}_j; \boldsymbol{u}_K) \in \{\mathrm{H}_0, \mathrm{H}_1\}$. We will later introduce a novel class of conditional independence tests that is particularly tailored to the algorithm and applicable to non-Gaussian continuous data. If $T_\alpha(\boldsymbol{u}_i, \boldsymbol{u}_j; \boldsymbol{u}_K) = \mathrm{H}_0$, the edge $i - j$ is removed from $\mathcal{G}$ and the conditioning set K is stored in two variables S_{ij} and S_{ji} for later use. As a result of the first step, $\mathcal{G}$ is turned into the skeleton of $\mathcal{G}^*$. Step one is summarised in Algorithm 3.

In the second step, $\mathcal{G}$ is transformed into a CG by introducing a v-structure $i \to k \leftarrow j$ whenever i and j are non-adjacent, $k \in \mathrm{ad}(i) \cap \mathrm{ad}(j)$, and $k \notin S_{ij}$. In the last step, $\mathcal{G}$ is transformed into $\mathcal{G}^*$ by directing further edges of $\mathcal{G}$ to prevent new v-structures and directed cycles, until no more edges need direction. Steps two and three are given in Algorithm 4, where the third step was taken from Pearl (2009, Section 2.5). If P is faithful to $\mathcal{D}$ and if all statistical test decisions made in Algorithm 3 are correct, then Algorithm 4 will return the correct graph $\mathcal{D}^e$, see Meek (1995). Due to the finite sample size or the existence of hidden variables, the application of Algorithm 3 to empirical data may sometimes, however, lead to conflicting information about edge directions. That is, it may be possible in a given situation that Algorithm 4, while introducing v-structures, first orients an undirected edge $i - j$ into $i \to j$, and later tries to introduce the opposite edge direction $i \leftarrow j$. In such a situation, we keep $i \to j$ and skip the new v-structure including $i \leftarrow j$. We can test whether the resulting CG can still be extended to a DAG without introducing new v-structures or directed cycles using the algorithm by Dor and Tarsi (1992). The PC algorithm can also be adapted to incorporate existing expert knowledge, see Meek (1995) and Moole and Valtorta (2004). We will henceforth assume that P is faithful to $\mathcal{D}$ and that there are no hidden variables.

4.1.2 Testing conditional independence using partial correlations

The centrepiece of the PC algorithm—as of any constraint-based estimation algorithm—is the test for conditional independence. In a Gaussian framework, the test

Algorithm 3 PC algorithm: finding the skeleton.

Input Data set $\boldsymbol{u}$; significance level $\alpha \in (0,1)$; conditional independence test with test decision $T_\alpha(\boldsymbol{u}_i, \boldsymbol{u}_j; \boldsymbol{u}_K)$ for $\mathrm{H}_0\colon U_i \perp\!\!\!\perp U_j \mid \boldsymbol{U}_K$, $i \neq j \in V$, $K \subseteq V \setminus \{i,j\}$.
Output Skeleton $\mathcal{G} = (V, E_\mathcal{G})$; separation sets S_{ij}, $i \neq j \in V$, $(i,j) \notin E_\mathcal{G}, (j,i) \notin E_\mathcal{G}$.

$\mathcal{G} \leftarrow$ complete UG on V;
$k \leftarrow 0$;
repeat
  for $i \in V$ **and** $j \in \mathrm{ad}(i)$ **do** % i and j are adjacent in $\mathcal{G}$
    if $T_\alpha(\boldsymbol{u}_i, \boldsymbol{u}_j; \boldsymbol{u}_K) = \mathrm{H}_0$ for any $K \subseteq \mathrm{ad}(i) \setminus \{j\}$ with $|K| = k$ **then**
      delete $i - j$ from $\mathcal{G}$;
      $S_{ij} \leftarrow K$;
      $S_{ji} \leftarrow K$;
    end if
  end for
  $k \leftarrow k + 1$.
until $|\mathrm{ad}(i)| \leq k$ for all $i \in V$.

Algorithm 4 PC algorithm: introducing edge directions.

Input Skeleton $\mathcal{G} = (V, E_\mathcal{G})$; separation sets S_{ij}, $i \neq j \in V$, $(i,j) \notin E_\mathcal{G}, (j,i) \notin E_\mathcal{G}$.
Output Chain graph $\mathcal{G}$.

% Introduce v-structures:
for $i \in V$ **and** $j \notin \mathrm{ad}(i)$ **and** $k \in \mathrm{ad}(i) \cap \mathrm{ad}(j)$ **do**
  if $k \notin S_{ij}$ **then**
    replace $i - k - j$ by $i \to k \leftarrow j$ in $\mathcal{G}$;
  end if
end for
% Orient as many undirected edges as possible by repeated application of the following rules:
repeat
  R1 orient $j - k$ into $j \to k$ whenever $\mathcal{G}$ contains $i \to j$ and $k \notin \mathrm{ad}(i)$;
  R2 orient $i - j$ into $i \to j$ whenever $\mathcal{G}$ contains $i \to k \to j$;
  R3 orient $i - j$ into $i \to j$ whenever $\mathcal{G}$ contains $i - k \to j$ and $i - l \to j$, and $l \notin \mathrm{ad}(k)$;
until no more edges can be directed;

of choice is usually a test for zero partial correlation $\rho_{ij\cdot K}$, see, for instance, Anderson (2003, Section 4.3). The null hypothesis then translates into $\mathrm{H}_0\colon \rho_{ij\cdot K}(X_i, X_j; X_K) = 0$, where $X_k \coloneqq \Phi^{-1}(U_k)$ for all $k \in V$, and Φ denotes the univariate standard normal cdf. Here, the quantile function Φ^{-1} is applied to U in order to transform the uniform univariate copula margins to standard normal margins. The conditional independence test is based on the asymptotic normality

$$\sqrt{d - |K| - 3}\, \widehat{z}_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathrm{N}(0,1), \qquad \widehat{z}_n \coloneqq \frac{1}{2}\log\left(\frac{1+\widehat{\rho}_{ij\cdot K}(\boldsymbol{X}_i^n, \boldsymbol{X}_j^n; \boldsymbol{X}_K^n)}{1-\widehat{\rho}_{ij\cdot K}(\boldsymbol{X}_i^n, \boldsymbol{X}_j^n; \boldsymbol{X}_K^n)}\right),$$

of the Fisher's z-transformed partial-correlation estimator $\widehat{\rho}_{ij\cdot K}$ under H_0, see again Anderson (2003, Section 4.3). Here, $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution, $\mathrm{N}(0,1)$ is the univariate standard normal distribution, and $\boldsymbol{X}_k^n \coloneqq (\Phi^{-1}(U_k^1), \dots, \Phi^{-1}(U_k^n))$ for all $k \in V$. Kalisch and Bühlmann (2007) have proven uniform convergence of the PC algorithm under joint normality and a mild sparsity assumption for the underlying DAG, cf. also Harris and Drton (2012). An implementation of the PC algorithm with above partial correlation test is available in the R package `pcalg` (Kalisch et al., 2012). The `pcalg` package also provides an interface for self-implemented conditional independence tests.

4.1.3 Vine-copula-based conditional independence test

Above test for zero partial correlation was derived under the assumption of joint normality. We now introduce a copula-based alternative test for conditional independence that is also applicable to non-Gaussian continuous data. Assume $K \neq \emptyset$. Otherwise, the problem reduces to testing ordinary (unconditional) stochastic independence. Let $F_{i,j|K}(\,\cdot\,,\cdot \,|\, \boldsymbol{v}_K)$ denote the conditional cdf of U_i and U_j given $\boldsymbol{U}_K = \boldsymbol{v}_K$, and let $C_{i,j|K}(\,\cdot\,,\cdot \,|\, \boldsymbol{v}_K)$ be the corresponding conditional copula. Moreover, let $C^{\perp\!\!\!\perp}$ again denote the independence copula on $[0,1]^2$. The conditional independence $U_i \perp\!\!\!\perp U_j \mid \boldsymbol{U}_K$ holds if and only if

$$\begin{aligned} F_{i,j|K}(v_i, v_j \,|\, \boldsymbol{v}_K) &= C_{i,j|K}\big(F_{i|K}(v_i \,|\, \boldsymbol{v}_K), F_{j|K}(v_j \,|\, \boldsymbol{v}_K) \,\big|\, \boldsymbol{v}_K\big) \\ &= F_{i|K}(v_i \,|\, \boldsymbol{v}_K)\, F_{j|K}(v_j \,|\, \boldsymbol{v}_K) \end{aligned}$$

for all $v_i, v_j \in [0,1]$ and P_K-almost all $\boldsymbol{v}_K \in [0,1]^{|K|}$, where $\boldsymbol{U}_K \sim P_K$. Hence, the null hypothesis of the conditional independence test can be stated as

$$H_0 \colon C_{i,j|K}(\,\cdot\,,\cdot \,|\, \boldsymbol{v}_K) = C^{\perp\!\!\!\perp}(\,\cdot\,,\cdot\,) \text{ for } P_K\text{-almost all } \boldsymbol{v}_K \in [0,1]^{|K|}.$$

Using again the simplifying assumption that $C_{i,j|K}(\,\cdot\,,\cdot \,|\, \boldsymbol{v}_K)$ depends on $\boldsymbol{v}_K$ only through $F_{i|K}(\,\cdot \,|\, \boldsymbol{v}_K)$ and $F_{j|K}(\,\cdot \,|\, \boldsymbol{v}_K)$, we drop $\boldsymbol{v}_K$ from $C_{i,j|K}(\,\cdot\,,\cdot \,|\, \boldsymbol{v}_K)$ and approximate H_0 by the more accessible null hypothesis $H_0^* \colon C_{i,j|K}(\,\cdot\,,\cdot\,) = C^{\perp\!\!\!\perp}(\,\cdot\,,\cdot\,)$. The new null hypothesis H_0^* can be tested using any test for ordinary (unconditional) stochastic independence of two continuous random variables applied to the transformed observations $W_{i|K}^1, \ldots, W_{i|K}^n$ and $W_{j|K}^1, \ldots, W_{j|K}^n$, where

$$W_{i|K}^k \coloneqq F_{i|K}\big(U_i^k \,\big|\, \boldsymbol{U}_K^k\big) \quad \text{and} \quad W_{j|K}^k \coloneqq F_{j|K}\big(U_j^k \,\big|\, \boldsymbol{U}_K^k\big), \quad k \in \{1, \ldots, n\}. \tag{4.1}$$

Song (2009) called Equation (4.1) the *Rosenblatt transform* after Rosenblatt (1952), while Bergsma (2011) called it the *partial copula transform*. Given a realisation $\boldsymbol{u}$ of $\big(\boldsymbol{U}^1, \ldots, \boldsymbol{U}^n\big)$, the difficulty of this approach lies in the computation of the transformed realisations $\boldsymbol{w}_{i|K}$ and $\boldsymbol{w}_{j|K}$, where $w_{i|K}^k \coloneqq F_{i|K}\big(u_i^k \,\big|\, \boldsymbol{u}_K^k\big)$ and $w_{j|K}^k \coloneqq F_{j|K}\big(u_j^k \,\big|\, \boldsymbol{u}_K^k\big)$ for all $k \in \{1, \ldots, n\}$. Note that the conditional cdfs $F_{i|K}(\,\cdot \,|\, \boldsymbol{v}_K)$ and $F_{j|K}(\,\cdot \,|\, \boldsymbol{v}_K)$ are typically unknown and need to be estimated in the course of the testing procedure. Bergsma (2011) suggested the use of non-parametric kernel estimators for this task. By contrast, we propose a parametric estimation method that is based on vine copula models.

Estimating conditional cdfs using vine copula models

Taking another look at vine copula models, we observe that transformed realisations like $\boldsymbol{w}_{i|K}$ and $\boldsymbol{w}_{j|K}$ naturally emerge in the log-likelihood function. In fact, given any distinct $i, j \in V$ and $K \subseteq V \setminus \{i, j\}$, it is always possible to construct a regular vine $\mathcal{V} = (T_1, \ldots, T_p)$, $p \coloneqq 1 + |K|$, in which tree T_1 has vertex set $V_1 = \{i\} \cup \{j\} \cup K$ and tree T_p is of the form $i, l | K_{-l} \overset{i,j|K}{\text{---}} j, m | K_{-m}$ for some $l, m \in K$. The corresponding log-likelihood function $l\big(\boldsymbol{\theta}; \boldsymbol{u}_{\{i\}\cup\{j\}\cup K}\big)$, $\boldsymbol{\theta} \in \Theta$, contains the pair-copula pdf $c_{i,j|K}$

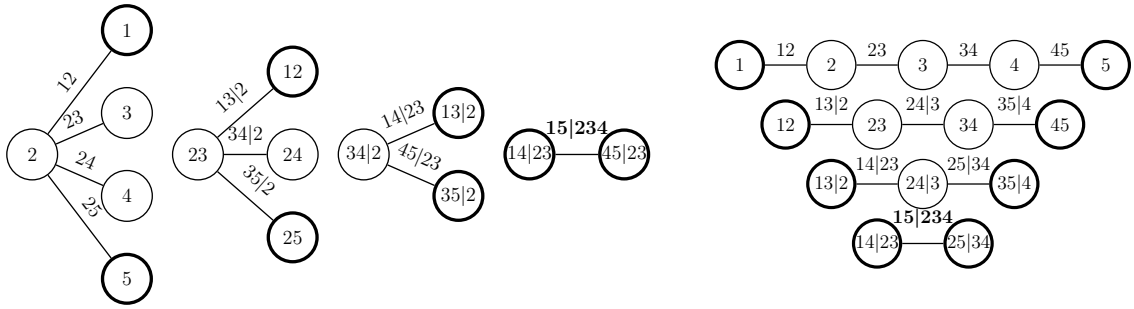


Figure 4.1: A C- (left) and a D-vine (right) on five vertices having the same edge label 15|234 in tree T_4. A corresponding R-vine is given in Figure 2.2. The three vines were constructed according to rule **R** with $i = 1$ and $j = 5$. Boundaries of nodes including either 1 or 5 appear in bold.

with arguments $F_{i|K}\big(u_i^k \,\big|\, \boldsymbol{u}_K^k; \boldsymbol{\theta}\big)$ and $F_{j|K}\big(u_j^k \,\big|\, \boldsymbol{u}_K^k; \boldsymbol{\theta}\big)$ for all $k \in \{1, \dots, n\}$. Thus, by computing an ML estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ and subsequently evaluating l at $\widehat{\boldsymbol{\theta}}$, we obtain estimates $\widehat{w}_{i|K}^k \coloneqq F_{i|K}\big(u_i^k \,\big|\, \boldsymbol{u}_K^k; \widehat{\boldsymbol{\theta}}\big)$ and $\widehat{w}_{j|K}^k \coloneqq F_{j|K}\big(u_j^k \,\big|\, \boldsymbol{u}_K^k; \widehat{\boldsymbol{\theta}}\big)$ of $w_{i|K}^k$ and $w_{j|K}^k$, respectively, as a welcome side effect.

We call a vertex v in Tree T_q, $q \in \{1, \dots, p-1\}$, an *inner vertex* if $|\mathrm{ad}(v)| \geq 2$. In order to construct such a vine $\mathcal{V}$, we have to follow one simple rule:

R Neither i nor j may be part of an inner vertex in trees $T_1, \dots, T_{p-1}$ of $\mathcal{V}$.

Following **R**, it is even possible to restrict the class of R-vines to C- or D-vines. The only inner vertices of a C-vine are the root vertices of the trees $T_1, \dots, T_{p-1}$. Thus, in a C-vine obeying **R**, i and j do not appear in the root vertices of the respective trees. Similarly, in a D-vine obeying **R**, i and j only appear in the boundary vertices of trees $T_1, \dots, T_{p-1}$. Figures 2.2 and 4.1 give an example of a C-, a D-, and an R-vine, respectively, having the same edge label in tree T_p.

The tree structure of $\mathcal{V}$ can be estimated from $\boldsymbol{u}_{\{i\}\cup\{j\}\cup K}$ by adapting the greedy search strategies described in Section 2.3 to the new constraint **R**. An optimal C-vine obeying **R** is found by restricting the sets of possible root vertices for trees $T_1, \dots, T_{p-1}$ to vertices containing neither i nor j, respectively. In order to find an optimal D-vine obeying **R**, the unconstrained TSP usually solved has to be replaced

by a constrained TSP with fixed source vertex i and destination vertex j. Finally, an optimal R-vine obeying $\mathbf{R}$ is found by first estimating a smaller R-vine $\mathcal{V}_K$ with first tree vertices K. Having found $\mathcal{V}_K$, vertex i is then connected to a vertex $l \in K$ in tree T_1 such that the new edge $i - l$ has optimal edge weight amongst all possible edges $i - m$ for $m \in K$. The same is done for vertex j. Note that this way, j cannot be connected to i. The newly formed structure is then sequentially transformed into $\mathcal{V}$ by analogously extending the remaining trees $T_2, \dots, T_p$, such that the proximity condition and $\mathbf{R}$ are always satisfied and the corresponding edge weights are optimised. Copula selection and ML estimation in the resulting vine copula model is then performed as usual.

Conditional independence tests

Summing up, we test the conditional independence $U_i \perp\!\!\!\perp U_j \mid \boldsymbol{U}_K$ in three steps. In the first step, we construct a vine $\mathcal{V}$ on the vertices $\{i\} \cup \{j\} \cup K$ by applying a modified version of one of the structure estimation algorithms described in Section 2.3 to $\boldsymbol{u}_{\{i\}\cup\{j\}\cup K}$. In the second step, we select corresponding pair-copula families, perform ML estimation in the resulting model, and evaluate the log-likelihood function l at the estimated parameter vector $\widehat{\boldsymbol{\theta}}$ to obtain transformed realisations $\widehat{\boldsymbol{w}}_{i|K} \coloneqq \left(\widehat{w}^k_{i|K}\right)_{1\le k\le n}$ and $\widehat{\boldsymbol{w}}_{j|K} \coloneqq \left(\widehat{w}^k_{j|K}\right)_{1\le k\le n}$, respectively. In the last step, we apply a test for ordinary stochastic independence of two continuous random variables to $\widehat{\boldsymbol{w}}_{i|K}$ and $\widehat{\boldsymbol{w}}_{j|K}$. Note that in the first iteration step of Algorithm 3, only unconditional independences, that is $K = \emptyset$, are tested, and thus the independence test of choice is directly applied to $\boldsymbol{u}$.

We examine the performance of our novel testing procedure in a simulation study in the next section, using three different tests for ordinary stochastic independence. Recycling notation, consider the null hypothesis $H_0 \colon U_i \perp\!\!\!\perp U_j$ vs. $H_1 \colon U_i \not\perp\!\!\!\perp U_j$. The first test used is a test for zero Kendall's τ with null hypothesis $H_0^* \colon \tau(U_i, U_j) = 0$ vs. $H_1^* \colon \tau(U_i, U_j) \neq 0$. Under H_0, the Kendall's τ estimator $\widehat{\tau}_n$ exhibits the asymptotic normality

$$\sqrt{\frac{9n(n-1)}{2(2n+5)}}\, \widehat{\tau}_n(\boldsymbol{U}_i, \boldsymbol{U}_j) \xrightarrow[n\to\infty]{\mathcal{L}} \mathrm{N}(0,1),$$

where $\boldsymbol{U}_i := \left(U_i^1, \ldots, U_i^n\right)$ and $\boldsymbol{U}_j := \left(U_j^1, \ldots, U_j^n\right)$, see Hollander and Wolfe (1999, Section 8.1). In general, $\tau(U_i, U_j) = 0$ does not imply $U_i \perp\!\!\!\perp U_j$. However, for many popular copula families like the Clayton, the Gaussian, and the Gumbel copula families, H_0 and H_0^* are equivalent. The family of Student's t copulas serves as a counterexample. We then consider H_0^* an approximation for H_0. The other two independence tests used in Section 4.2 are of Cramér-von Mises type. More precisely, independence test number two is the test for zero Hoeffding's D proposed by Hoeffding (1948). P-values of the sample test statistic $\widehat{D}_n$ are computed using the asymptotically equivalent sample test statistic $\widehat{B}_n$ by Blum et al. (1961), see also Hollander and Wolfe (1999, Section 8.6). Independence test number three is the test by Genest and Rémillard (2004) based on the empirical copula process.

4.2 Structure estimation: A simulation study

We conducted an extensive simulation study to examine the small sample performance of the PC algorithm in finding the true Markov structure underlying a PCBN.

4.2.1 Simulation setup

We drew samples from various PCBNs based on the conditional independence properties represented by the DAG $\mathcal{D}$ in Figure 1.2. These PCBNs emerged from various choices of pair-copula families for C_{21}, C_{31}, C_{42}, and $C_{43|2}$. More precisely, we chose from the Clayton, Gumbel, Gaussian, and Student's t pair-copula families. These copula families exhibit considerable differences in their dependence structures and tail behaviours. We considered four PCBNs with all four pair copulas C_{21}, C_{31}, C_{42}, and $C_{43|2}$ coming from the same copula family, respectively. Additionally, we considered 24 PCBNs with each pair copula C_{21}, C_{31}, C_{42}, and $C_{43|2}$ coming from a different copula family. Our choices of pair-copula families are given in Table 4.1. For each choice of pair-copula families we then considered 16 different parameter configurations arising from a selection of two different parameter values for each pair copula. The parameter values for each pair copula were chosen to correspond

to values of Kendall's τ of 0.25 and 0.75, that is one low and one high rank correlation specification. These configurations are summarised in Table 4.2.

Our selection of copula parameters is based on the bijective relationship between the parameters of the Clayton, Gumbel, and Gaussian pair-copula families and the corresponding Kendall's τ. For the Student's t copula, such a bijective relationship exists only between the correlation parameter and Kendall's τ, which is why we set the degrees-of-freedom parameter of each Student's t copula to $\nu = 5$ in order to allow for heavy-tailed dependence. See Table 3.3 for the relations between parameters, TDCs, and Kendall's τ for each pair copula used in the simulation study.

Summing up, we have 28 different PCBNs with 16 different parameter configurations each, that is, 448 simulation scenarios. In each of the 448 simulation scenarios we performed $N = 100$ simulation runs, and in each simulation run we generated $n = 1{,}000$ i.i.d. observations. For each of the 44,800 runs we applied the PC algorithm with the ten different conditional independence tests described in Section 4.1.3. Those were the widely used test for zero partial correlation (COR) and our novel vine-copula-based tests using either only C-vines (C), or only D-vines (D), or more generally R-vines (R), respectively, together with one of the Kendall's τ (K), Hoeffding's D (H), or Genest and Rémillard (GR) tests for ordinary (unconditional) stochastic independence. Since zero partial correlation is generally not equivalent to conditional independence, we consider COR only an approximate conditional independence test serving as a benchmark. In a Gaussian framework, however, zero partial correlation is equivalent to conditional independence. This equivalence holds in particular in the scenarios featuring only Gaussian pair copulas, in which case the respective joint copula families are also Gaussian. The corresponding correlation matrices were derived in Section 3.4. Each test was performed at the 5% level.

4.2.2 Results

Let $\mathcal{G}_{f,p,r,t}$ denote the CG obtained from applying the PC algorithm with conditional independence test $t \in \{\text{COR}, \text{C-GR}, \text{C-H}, \text{C-K}, \text{D-GR}, \text{D-H}, \text{D-K}, \text{R-GR}, \text{R-H}, \text{R-K}\}$ to the data simulated in run $r \in \{1, \ldots, 100\}$ of pair-copula scenario $f \in \{1, \ldots, 28\}$

Copula	1	2	3	4	5	6	7	8	9	10	11	12	13	14
C_{21}	C	G	N	t	C	C	C	C	C	C	G	G	G	G
C_{31}	C	G	N	t	G	G	N	N	t	t	C	C	N	N
C_{42}	C	G	N	t	N	t	G	t	G	N	N	t	C	t
$C_{43\|2}$	C	G	N	t	t	N	t	G	N	G	t	N	t	C
Copula	15	16	17	18	19	20	21	22	23	24	25	26	27	28
C_{21}	G	G	N	N	N	N	N	N	t	t	t	t	t	t
C_{31}	t	t	C	C	G	G	t	t	C	C	G	G	N	N
C_{42}	C	N	G	t	C	t	C	G	G	N	C	N	C	G
$C_{43\|2}$	N	C	t	G	t	C	G	C	N	G	N	C	G	C

Table 4.1: Selected pair-copula families for C_{21}, C_{31}, C_{42}, $C_{43|2}$. Copulas were chosen from the Clayton (C), Gumbel (G), Gaussian (N), and Student's t (t) pair-copula families. See Tables 4.2 and 3.3 for further details on the pair-copula families used.

Copula	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
C_{21}	0.25	0.75	0.25	0.25	0.25	0.75	0.75	0.75	0.25	0.25	0.25	0.75	0.75	0.75	0.25	0.75
C_{31}	0.25	0.25	0.75	0.25	0.25	0.75	0.25	0.25	0.75	0.75	0.25	0.75	0.75	0.25	0.75	0.75
C_{42}	0.25	0.25	0.25	0.75	0.25	0.25	0.75	0.25	0.75	0.25	0.75	0.75	0.25	0.75	0.75	0.75
$C_{43\|2}$	0.25	0.25	0.25	0.25	0.75	0.25	0.25	0.75	0.25	0.75	0.75	0.25	0.75	0.75	0.75	0.75

Table 4.2: Selected values of Kendall's τ for each choice of pair-copula families for C_{21}, C_{31}, C_{42}, $C_{43|2}$. See Tables 4.1 and 3.3 for further details on the pair-copula families used.

(see Table 4.1) and parameter configuration $p \in \{1, \dots, 16\}$ (see Tables 4.2 and 3.3). We compared each CG $\mathcal{G}_{f,p,r,t}$ to the true essential graph $\mathcal{D}^e$ in Figure 1.2, and set $\pi_{f,p,r,t} := 1$ if $\mathcal{G}_{f,p,r,t}$ equalled $\mathcal{D}^e$ and $\pi_{f,p,r,t} := 0$ otherwise. For each pair-copula scenario f and each conditional independence test t, we then computed the relative frequency of recovering the correct structure over all parameter configurations p and all runs r, which we will denote by $\pi_{f,t} := \frac{1}{1600} \sum_{p=1}^{16} \sum_{r=1}^{100} \pi_{f,p,r,t}$. Moreover, we determined the *structural Hamming distance* (SHD) (Tsamardinos et al., 2006) $\delta_{f,p,r,t}$ between each CG $\mathcal{G}_{f,p,r,t}$ and $\mathcal{D}^e$. In short, $\delta_{f,p,r,t}$ counts the number of edges that need to be added to, removed from, directed in, or flipped in $\mathcal{G}_{f,p,r,t}$ in order to obtain $\mathcal{D}^e$. Hence, $\delta_{f,p,r,t}$ takes a value between zero and $\binom{|V|}{2} = 6$. We again took the average over all parameter configurations p and all runs r, yielding the mean SHD $\delta_{f,t} := \frac{1}{1600} \sum_{p=1}^{16} \sum_{r=1}^{100} \delta_{f,p,r,t}$ for each pair-copula scenario f and each conditional independence test t. The results are given in Figures 4.2 and 4.3, respectively.

Let us first consider Figure 4.2. The relative frequencies $\pi_{f,\text{COR}}$ range between 14% and 63%, whereas for the vine-copula-based tests, $\pi_{f,t}$ ranges between 40% and 64%. COR was outperformed by at least one vine-copula-based test in 18, and by all vine-copula-based tests in 15 out of the 28 copula scenarios. The lowest frequency of 14% was obtained when applying the PC algorithm with COR to the data sets generated in copula scenario 1 (numbering as in Table 4.1), which features only Clayton, that is non-elliptical, copulas. By contrast, COR showed a solid performance in the elliptical-copulas-only scenarios 3 and 4, which is not surprising given that COR is based on the partial correlation. In 9 out of the 28 copula scenarios, $\pi_{f,\text{COR}}$ is lower than 40%, which is the minimum frequency obtained for the vine-copula-based tests. Also, in these 9 scenarios, the difference in relative frequencies between COR and the vine-copula-based tests ranges between 9 and 33 percentage points. The highest frequency of 64% was obtained in copula scenario 15 both for the PC algorithm with C-GR and C-H, respectively. Taking means over all 28 copula scenarios, we obtain the overall relative frequencies $\pi_t := \frac{1}{28} \sum_{f=1}^{28} \pi_{f,t}$ for all tests t. These overall frequencies range between 50% and 53% for the vine-copula-based tests, while $\pi_{\text{COR}} = 45\%$. The best performances were again achieved by C-GR and C-H. However, we recommend using the R-vine-based conditional independence

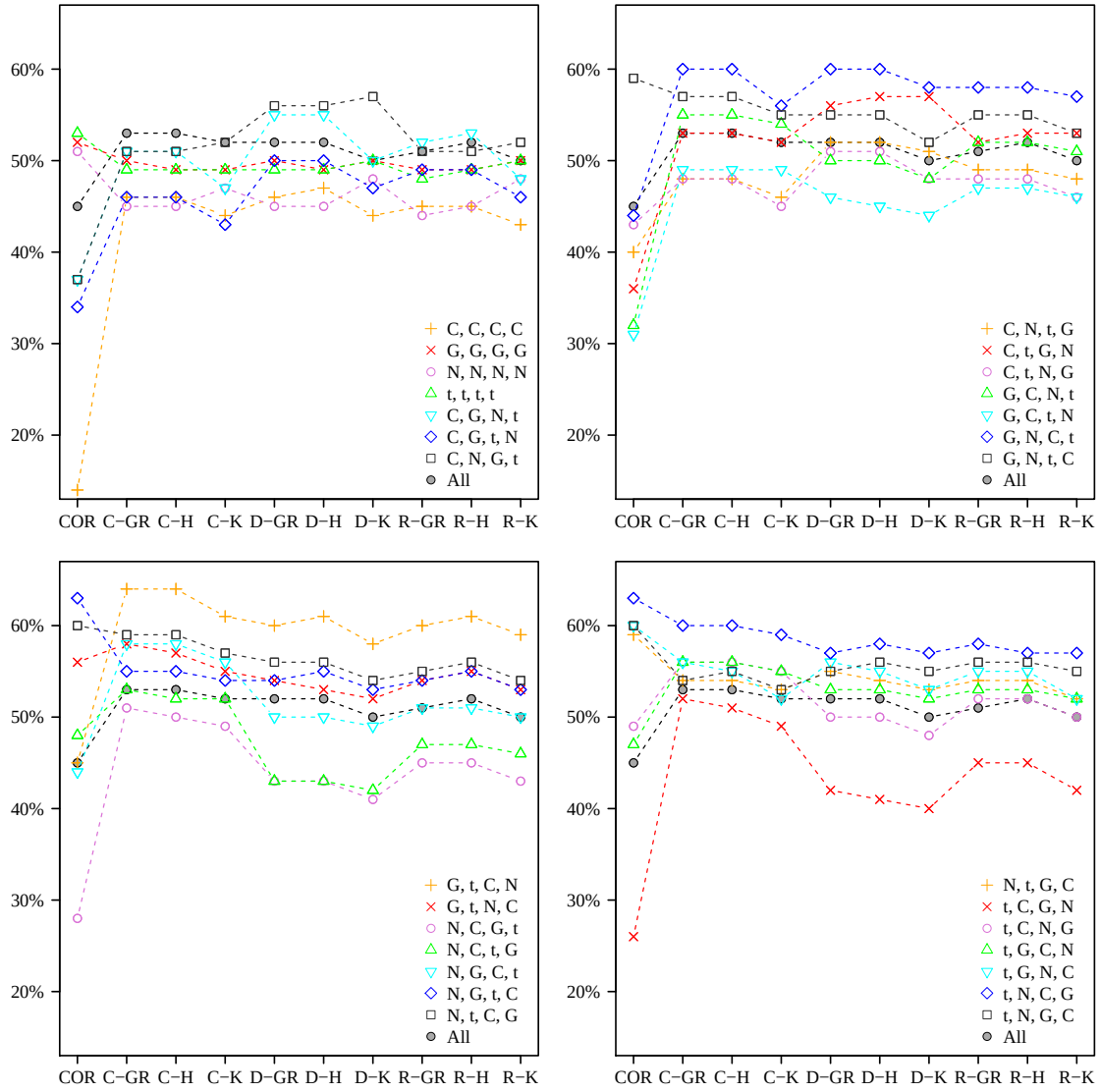


Figure 4.2: Percentage $\pi_{f,t}$ of runs in which the PC algorithm returned the correct Markov structure for each choice f of pair-copula families for C_{21}, C_{31}, C_{42}, $C_{43|2}$ (legends) and each independence test t (horizontal axes) (1,600 runs each). Copulas were chosen from the Clayton (C), Gumbel (G), Gaussian (N), and Student's t (t) pair-copula families. The percentage of correct recoveries out of all 28 copula scenarios is given in solid grey.

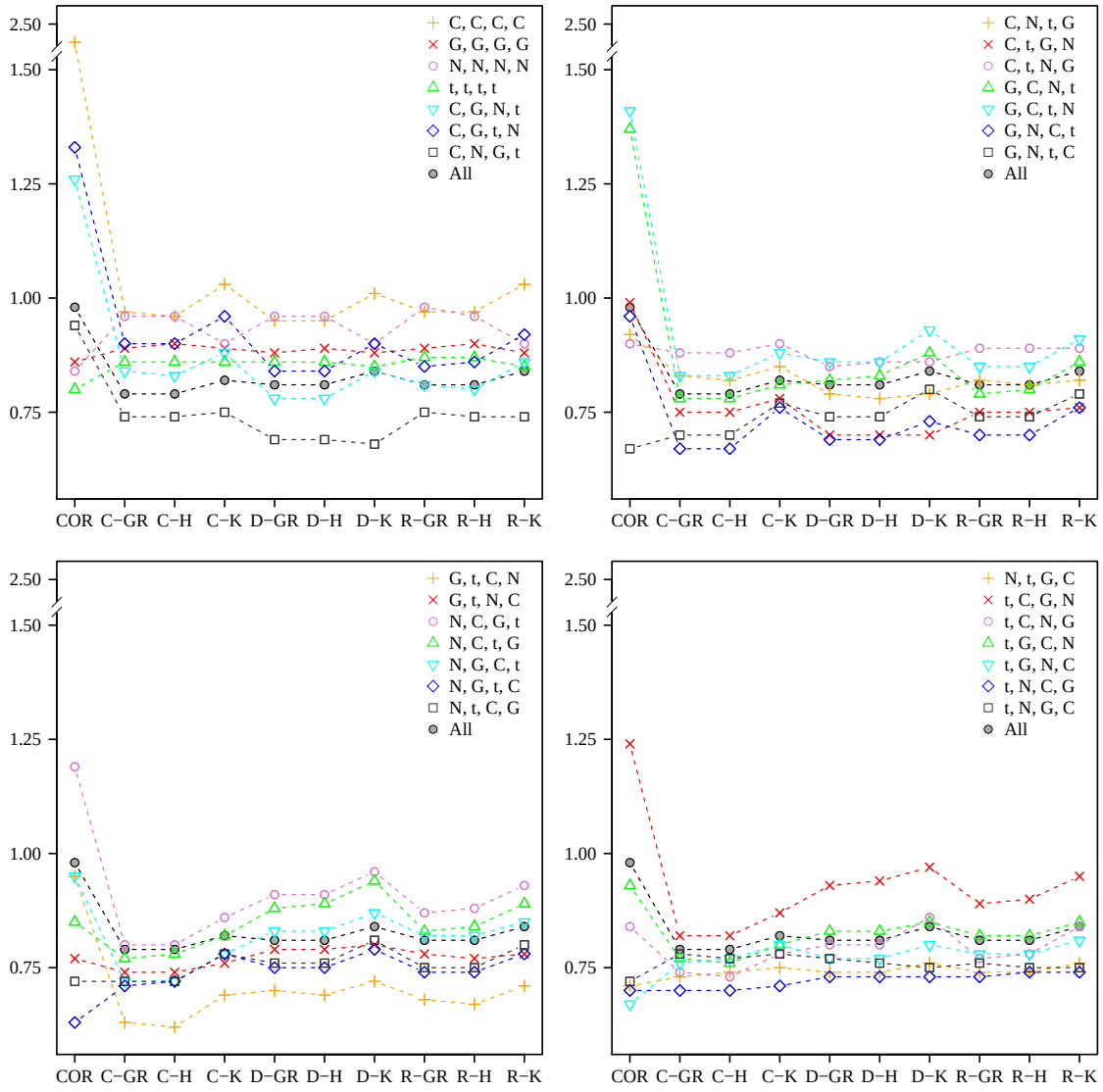


Figure 4.3: Average structural Hamming distance (SHD) $\delta_{f,t}$ between the true essential graph $\mathcal{D}^e$ and the CG $\mathcal{G}_{f,p,r,t}$ returned by the PC algorithm for each choice f of pair-copula families for C_{21}, C_{31}, C_{42}, $C_{43|2}$ (legends) and each independence test t (horizontal axes) (1,600 runs each). Copulas were chosen from the Clayton (C), Gumbel (G), Gaussian (N), and Student's t (t) pair-copula families. The average SHD over all 28 copula scenarios is given in solid grey.

tests in higher dimensions since these offer more general tree structures than their C- and D-vine counterparts. Moreover, we observe that choosing H instead of GR as test for unconditional stochastic independence has only little effect on the performance of the vine-copula-based tests. By contrast, relative frequencies were, on average, slightly worse when using K instead of GR and H, respectively. Since zero Kendall's τ is generally also not equivalent to stochastic independence, we recommend using GR and H. Note that in a given copula scenario f and a given parameter scenario p, the relative frequencies $\pi_{f,p,t} := \frac{1}{100}\sum_{r=1}^{100}\pi_{f,p,r,t}$ can be a lot higher than the averages displayed in Figure 4.2. We observed frequencies $\pi_{f,p,t}$ of up to 98%. To sum up, using a vine-copula-based conditional independence test instead of COR leads to more reliable structure estimates, in particular when the data exhibit non-Gaussian, asymmetric dependence.

Considering only the correctly recovered Markov structures may be a too crude performance measure. Hence, the mean SHDs $\delta_{f,t}$ in Figure 4.3 illustrate how much the results of the PC algorithm differ from the true essential graph $\mathcal{D}^e$. For the vine-copula-based tests, $\delta_{f,t}$ ranges between 0.62 and 1.03. The respective overall means $\delta_t := \frac{1}{28}\sum_{f=1}^{28}\delta_{f,t}$ lie between 0.79 and 0.84. Thus, on average, the results of the PC algorithm differ by less than one edge from $\mathcal{D}^e$. That is, if the PC algorithm yields a CG that is not equivalent to $\mathcal{D}^e$, then, with a high probability, CG and $\mathcal{D}^e$ are not too different. The lowest values of δ_t were again obtained for C-GR and C-H. Similarly, $\delta_{f,\mathrm{COR}}$ ranges between 0.63 and 2.44, and $\delta_{\mathrm{COR}} = 0.98$, which again shows the superiority of the vine copula approach. The worst mean SHD of 2.44 was obtained in copula scenario 1. Overall, we can say that the PC algorithm with either of the 9 vine-copula-based conditional independence tests provides a suitable procedure for structure estimation in PCBNs.

We repeated the simulation study both for a significance level α of 1% and for a sample size n of 500. For $\alpha = 1\%$, we obtained results similar to the ones described above for $\alpha = 5\%$. The overall relative frequencies π_t were slightly lower, ranging from 44% to 47% for the vine-copula-based tests, while π_{COR} was 43%. Also, the overall mean SHDs δ_t ranged between 0.86 and 0.94 for the vine-copula-based tests,

while π_{COR} was 0.99. The reduction in sample size to $n = 500$, on the other hand, lead to a slightly stronger decrease in the overall relative frequencies π_t, which then ranged between 39% and 41% for the vine-copula-based tests, while π_{COR} was 37%. Similarly, the overall mean SHDs δ_t ranged between 1.07 and 1.11 for the vine-copula-based tests, while π_{COR} was 1.17. Yet, both for $\alpha = 1\%$ and for $n = 500$, the CGs returned by the PC algorithm differed on average from $\mathcal{D}^e$ by only one edge. The performance of the PC algorithm can thus be deemed reliable and robust. In future research, one may investigate the performance of other conditional independence tests like Zhang et al. (2011), as well as of other estimation algorithms.

4.3 Ordering the parents of a DAG

Given the DAG $\mathcal{D} = (V, E)$ and a set $\mathcal{O}$ of parent orderings, the selection of pair-copula families for $C_{v,w|\text{pa}(v;w)}$, $v \in V$, $w \in \text{pa}(v)$, can be performed in a similar way as in Section 2.3 for vine copula models, with the difference that the iteration is vertex-by-vertex and parent-by-parent instead of tree-by-tree.

For the selection of $\mathcal{O}$ we propose a greedy-type procedure inspired by the structure selection algorithm for vine copula models outlined in Section 2.3. Clearly, an ordering of the parents of a vertex $v \in V$ is only required if $\text{pa}(v) \neq \emptyset$. We assume that $\mathcal{D}$ is well-ordered. Let $v \in V$ and assume $k := |\text{pa}(v)| \geq 1$. Moreover, let $i \in \{1, \dots, k\}$ and assume that we have already selected the $i-1$ smallest parents of v, denoted by $w_1 <_v \cdots <_v w_{i-1}$. This implies that we have already selected pair-copula families for $C_{v^*,w|\text{pa}(v^*;w)}$, v^* smaller than v by the well-ordering of $\mathcal{D}$, $w \in \text{pa}(v^*)$, and $C_{v,w_j|\text{pa}(v;w_j)}$, $j < i$. Also, this implies that we have inferred corresponding ML parameter estimates, which we summarise in the vector $\widehat{\boldsymbol{\theta}}$. Let $W_{-i} := \{w_1, \dots, w_{i-1}\}$. The selection of w_i is performed in three steps. First, we compute the pseudo-observations $F_{v|W_{-i}}\big(u_v^k \,\big|\, \boldsymbol{u}_{W_{-i}}^k; \widehat{\boldsymbol{\theta}}\big)$ and $F_{w|W_{-i}}\big(u_w^k \,\big|\, \boldsymbol{u}_{W_{-i}}^k; \widehat{\boldsymbol{\theta}}\big)$, $k \in \{1, \dots, n\}$, for all $w \in \text{pa}(v) \setminus W_{-i}$. Note that for $i = 1$, nothing needs to be done since all univariate marginals are uniform on $[0, 1]$. Second, we assign a weight $\omega_{v,w}$ to every edge $w \to v$, $w \in \text{pa}(v) \setminus W_{-i}$, based on the previously calculated pseudo-observations, and choose w_i such that $w_i \to v$ has optimal edge weight.

Suitable weights are, for instance, the absolute values of estimates of Kendall's τ, or AIC or BIC values of selected pair-copula families with estimated parameters. Last, we select a pair-copula family for $C_{v,w_i|\mathrm{pa}(v;w_i)}$ and compute an ML estimate of the corresponding parameter(s). Again, this last step may have already been performed when computing the edge weights $\omega_{v,w}$.

5 Data applications

PCBNs can accommodate a great variety of distributional features to be modelled such as heavy-tailedness and non-linear, asymmetric dependence, while, at the same time, they can capture pre-specified Markov properties (as given by a DAG). In this chapter, the routines presented in earlier sections are applied to modelling financial return data. Moreover, we develop strategies for reducing the computational effort potentially involved in evaluating the log-likelihood of a PCBN.

5.1 German and US stock and bond market indices

As a real-world application, we applied PCBNs to a four-variate financial data set comprising US and German stock and bond market indices. More precisely, we modelled the dependence structure of daily log-returns of the Dow Jones Industrial Average (DJI), the Dow Jones Corporate Bond Index (DJCB), the German stock index (DAX), and the corresponding German corporate bond index (RDAX) from 3 April 2007 to 30 September 2010 ($n = 854$ observations). US indices are given in US Dollars and German indices in Euros, that is, we did not correct the data for exchange rate fluctuations. Figure 5.1 shows the four time series of daily log-returns.

5.1.1 Univariate time series models

Using the inference functions for margins method outlined in Section 2.3, we modelled univariate marginal distributions without regard to the dependence structure between variables. We first removed serial correlation in the four time series of log-returns by applying an AR(1)-GARCH(1,1) filter, which accounts for conditional

	μ [$\times 10^3$]	a	ω [$\times 10^5$]	α	β	ν
DJI	1.02 (0.14)	−0.10 (0.04)	0.20 (2.88)	0.12 (0.02)	0.88 (0.02)	6.33 (1.52)
DJCB	0.16 (0.37)	−0.11 (0.03)	0.06 (3.77)	0.08 (0.02)	0.89 (0.02)	6.75 (1.73)
DAX	0.58 (0.44)	0.01 (0.03)	0.39 (3.79)	0.09 (0.02)	0.89 (0.02)	8.24 (2.30)
RDAX	0.25 (0.07)	0.10 (0.04)	0.01 (2.38)	0.05 (0.01)	0.95 (0.01)	17.30 (9.03)

Table 5.1: ML estimates and standard errors (in parentheses) of AR(1)-GARCH(1,1) parameters for the DJI, DJCB, DAX, and RDAX daily log-returns.

heteroskedasticity present in the data, see Bollerslev (1986). The log-return $r_{i,t}$ of index $i \in \{\text{DJI}, \text{DJCB}, \text{DAX}, \text{RDAX}\}$ at time t can thus be written as

$$\begin{aligned} r_{t,i} &= \mu_i + a_i\, r_{t-1,i} + \varepsilon_{t,i}, \\ \varepsilon_{t,i} &= \sigma_{t,i}\, z_{t,i}, \\ \sigma^2_{t,i} &= \omega_i + \alpha_i\, \varepsilon^2_{t-1,i} + \beta_i\, \sigma^2_{t-1,i}, \end{aligned} \tag{5.1}$$

with parameters $\omega_i > 0$, $\alpha_i, \beta_i \geq 0$ such that $\alpha_i + \beta_i < 1$, $|a_i| < 1$, and $\mu_i \in \mathbb{R}$, where $\mathbb{E}[z_{t,i}] = 0$ and $\mathbb{V}\text{ar}[z_{t,i}] = 1$. The standardised residuals $z_{t,i}$ are assumed to follow a univariate Student's t distribution with ν_i degrees of freedom, that is, $\sqrt{\frac{\nu_i}{\nu_i - 2}}\, z_{t,i} \sim \text{t}_{\nu_i}$. ML parameter estimates and their standard errors derived from numerical evaluation of the Hessian of the AR(1)-GARCH(1,1) parameters are given in Table 5.1. Using these standard errors and a 5% significance level, we cannot reject the null hypothesis of the Ljung-Box test that there is no autocorrelation left in the residuals and squared residuals (Ljung and Box, 1978). The same holds true for the null hypothesis of the Lagrange-multiplier ARCH test that the residuals exhibit no conditional heteroskedasticity (Engle, 1982). We converted the standardised residuals to uniformly distributed observations $u_{t,i} = \text{t}_{\nu_i}\left(\sqrt{\frac{\nu_i}{\nu_i - 2}}\, z_{t,i}\right)$ before modelling the joint dependence structure of the four time series of log-returns by a PCBN.

5.1.2 Specifying the conditional independence structure

Based on the economic consideration that the German stock index is driven by its US counterpart and that within the US and Germany corporate bond indices are

driven by the respective national stock indices, we propose a conditional independence model for the transformed residuals $u_{t,i}$. The DAG $\mathcal{D}$ from Figure 1.2 with vertices 1, 2, 3, 4 representing the variables DJI, DJCB, DAX, RDAX, respectively, reflects the above-mentioned dependences and specifies the conditional independence assumptions

$$\text{DJCB} \perp\!\!\!\perp \text{DAX} \mid \text{DJI} \qquad \text{and} \qquad \text{DJI} \perp\!\!\!\perp \text{RDAX} \mid \{\text{DJCB}, \text{DAX}\}.$$

Besides, we also obtain the DAG $\mathcal{D}$ when applying the PC algorithm with either of the ten conditional independence tests COR, C-GR, C-H, C-K, D-GR, D-H, D-K, R-GR, R-H, and R-K described in Section 4.1.3 (with notation as in Section 4.2). Given the results presented below, we retrospectively measured the reliability of the PC algorithm for the analysed financial data set as follows. We first generated $N = 100$ i.i.d. samples of size $n = 854$ from the non-Gaussian PCBN specified by the joint ML parameter estimates in Table 5.2 and then applied the PC algorithm to recover the conditional independence properties of each of these samples. The true DAG structure was recovered in 88% to 93% of the cases (depending on the conditional independence test used), which supports our model assumptions.

Given the DAG $\mathcal{D}$, Theorem 3.1 prescribes which pair copulas need specification in the definition of our model. Note that vertex 4 (RDAX) has two parents (DJCB and DAX), which necessitates the selection of a parent ordering. Using the procedure described in Section 4.3 with Kendall's τ edge weights, we decided to use vertex 2 (DJCB) as the first parent. Our decision was based on estimates $\widehat{\tau}$ of Kendall's τ between vertices $2, 4$ ($\widehat{\tau} = 0.39$) and $3, 4$ ($\widehat{\tau} = -0.25$), respectively. The corresponding log-likelihood function was derived in Example 3.6.

5.1.3 Pair-copula selection and ML estimation

Having fixed our model's Markov structure, we next selected parametric copula families for C_{21}, C_{31}, C_{42}, and $C_{43|2}$, respectively. We considered the Clayton, Frank, Gaussian, Gumbel, and Student's t copula families as well as reflected versions of

the Clayton and Gumbel copula families in order to account for negative correlations. Among these candidates, we selected copula families based on comparisons of AIC values, which resulted in modelling all four copulas C_{21}, C_{31}, C_{42}, and $C_{43|2}$ by Student's t copulas. Figure 5.2 displays these choices along with kernel density estimates of the respective true copula pdfs. Visual comparison of these plots strongly affirms our choices of the Student's t copula family. We shall note that the estimates of the correlation parameters ρ and the degrees of freedom ν used to compute AIC values are nothing else than the sequential ML estimates to be obtained from the selected PCBN. Our choice of t copula is consistent with popular modelling approaches in the literature on statistical finance, see Ignatieva and Platen (2010).

We then computed joint ML estimates of the parameters of the so selected non-Gaussian PCBN. In view of reducing model complexity, we also applied a semiparametric ML estimator inspired by Hobæk Haff (2013). To this end, we replaced the integral in the log-likelihood function in Example 3.6 by a non-parametric conditional cdf estimator of $F_{3|2}(\,\cdot \,|\, u_{t,2})$ given by

$$\widehat{F}_{3|2}(u_{t,3} \,|\, u_{t,2}) = \frac{\sum_{s=1}^{n} \Phi\Big(\frac{\Phi^{-1}(u_{t,3}) - \Phi^{-1}(u_{s,3})}{h_3}\Big)\, \phi\Big(\frac{\Phi^{-1}(u_{s,2}) - \Phi^{-1}(u_{t,2})}{h_2}\Big)}{\sum_{s=1}^{n} \phi\Big(\frac{\Phi^{-1}(u_{s,2}) - \Phi^{-1}(u_{t,2})}{h_2}\Big)}, \quad t \in \{1, \dots, n\},$$

where ϕ denotes the standard normal pdf, and h_2, h_3 are normal reference rule-of-thumb bandwidths, see Li and Racine (2007, Section 6.2). We used transformed observations $\Phi^{-1}(u_{t,i})$ for the kernel smoothing to avoid boundary effects. The log-likelihood function of this model, which we will refer to as the *semiparametric non-Gaussian PCBN*, takes the form

$$\begin{aligned}\widehat{L}(\boldsymbol{\theta}; \boldsymbol{u}) = \sum_{t=1}^{n} &\log c_{21}(u_{t,2}, u_{t,1}; \boldsymbol{\theta}_{21}) + \log c_{31}(u_{t,3}, u_{t,1}; \boldsymbol{\theta}_{31}) + \log c_{42}(u_{t,4}, u_{t,2}; \boldsymbol{\theta}_{42}) \\ &+ \log c_{43|2}\big(h_{\underline{4}2}(u_{t,4}, u_{t,2}; \boldsymbol{\theta}_{42}), \widehat{F}_{3|2}(u_{t,3} \,|\, u_{t,2}); \boldsymbol{\theta}_{43|2}\big).\end{aligned}$$

ML estimation in this model is performed as in the fully parametric case. Last, we also applied the Gaussian PCBN from Section 3.4 to the data and compared

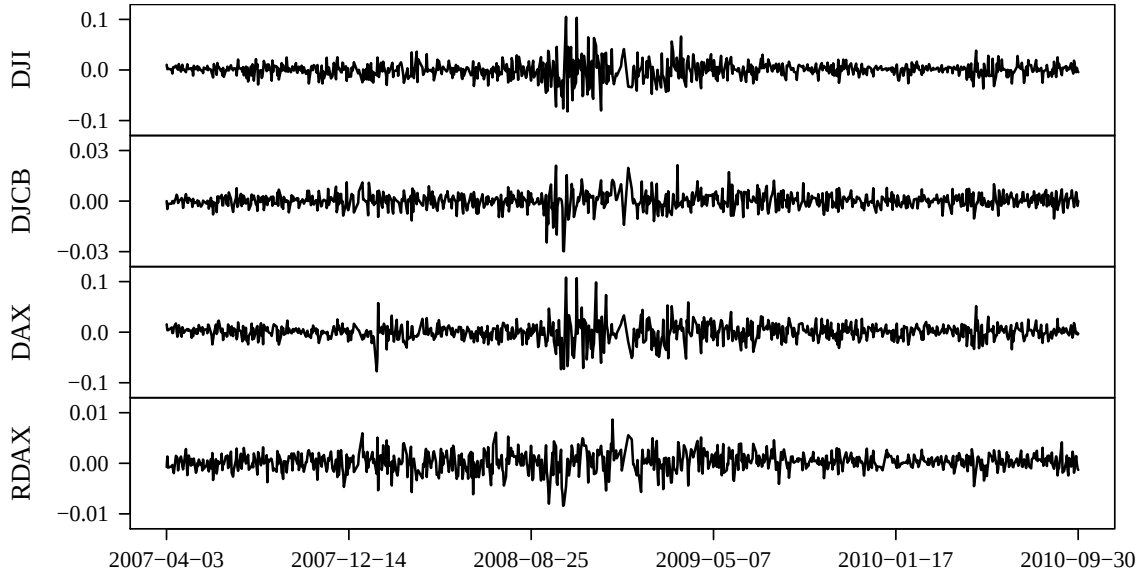
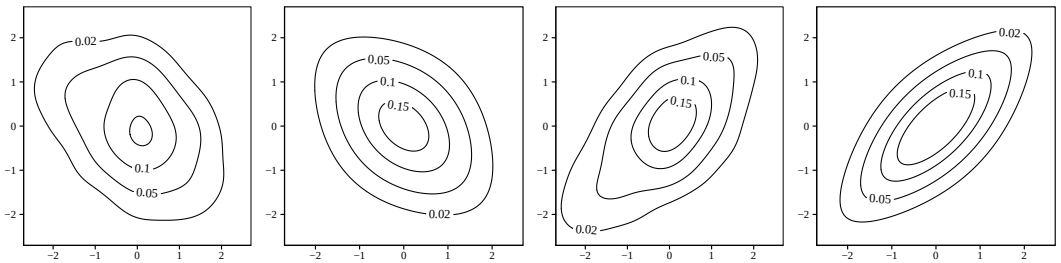
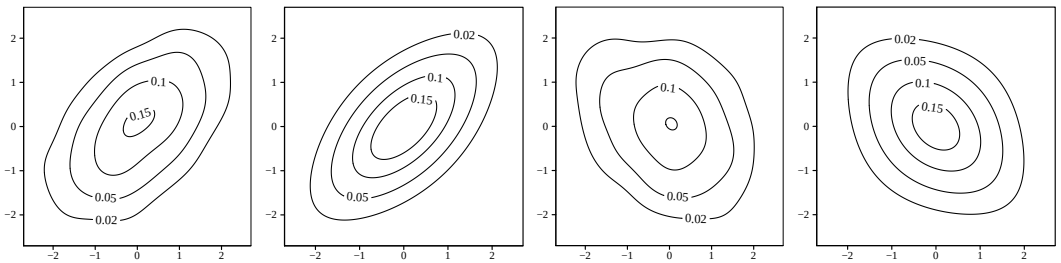


Figure 5.1: Daily log-returns of the Dow Jones Industrial Average (DJI), the Dow Jones Corporate Bond Index (DJCB), the German stock index (DAX), and the German corporate bond index (RDAX).



Copula pdf c_{21} (DJCB, DJI) Copula pdf c_{31} (DAX, DJI)



Copula pdf c_{42} (RDAX, DJCB) Copula pdf $c_{43|2}$ ({RDAX, DAX} | DJCB)

Figure 5.2: Kernel density estimates of pair-copula pdfs (left) and our choices of Student's t copula pdfs (right) for modelling the DJI, DJCB, DAX, and RDAX data. All copulas are displayed with standard normal margins.

its performance to the non-Gaussian and the semiparametric non-Gaussian PCBN. Parameter estimates, bootstrapped standard errors, and estimates of Kendall's τ for all three models are given in Table 5.2. The respective maximised log-likelihoods and AIC values are summarised in Table 5.3.

Due to non-normal tail behaviour observed in the data, it is not surprising that the Gaussian PCBN again turns out inferior to its non-Gaussian competitors. Applying the Vuong test with AIC correction (Vuong, 1989) for model selection yields the same conclusion. The results for the non-Gaussian and the semiparametric non-Gaussian PCBN, however, are rather close. In fact, the null hypothesis of the Vuong test that both models are equally close to the true model cannot be rejected at the 5% significance level. Choosing the semiparametric over the fully parametric non-Gaussian PCBN cut down computation time by the factor 50.

5.2 International stock market indices

We also applied PCBNs to a financial data set comprising ten major international stock market indices. More precisely, we modelled the joint distribution of a portfolio of daily log-returns of the Australian All Ordinaries (AUS), the Canadian S&P/TSX Composite Index (CAN), the Swiss Market Index (CH), the German DAX (DEU), the French CAC 40 (FRA), the Hong Kong Hang Seng Index (HK), the Japanese Nikkei 225 (JPN), the Singapore Straits Times Index (SGP), the UK's FTSE 100 (UK), and the US S&P 500 (USA) from 1 April 2008 to 29 July 2011 ($n = 733$ observations). The ten time series of daily log-returns are given in Figure 5.3.

5.2.1 Univariate time series models

Using the inference functions for margins method, we again modelled univariate marginal distributions without regard to the dependence structure between variables. We first removed serial correlation in the ten time series of log-returns by applying an AR(1)-GARCH(1,1) filter, as was given in Equation (5.1). The standardised residuals $z_{t,i}$ of stock index $i \in \{\text{AUS}, \text{CAN}, \text{CH}, \text{DEU}, \text{FRA}, \text{HK}, \text{JPN}, \text{SGP}, \text{UK}, \text{USA}\}$

Copula		non-Gaussian		semiparametric non-Gaussian		Gaussian	
		Parameters	$\widehat{\tau}$	Parameters	$\widehat{\tau}$	Parameters	$\widehat{\tau}$
C_{21}	S t	$-0.35, 10.1$ $(0.03, 3.8)$	-0.23	t $-0.35, 10.1$ $(0.03, 3.9)$	-0.23	N -0.34 (0.03)	-0.22
	J t	$-0.35, 10.4$ $(0.03, 4.3)$	-0.23	t $-0.35, 10.1$ $(0.03, 4.2)$	-0.23	N -0.34 (0.03)	-0.22
C_{31}	S t	$0.66, 9.2$ $(0.02, 4.3)$	0.46	t $0.66, 9.2$ $(0.02, 4.3)$	0.46	N 0.66 (0.02)	0.46
	J t	$0.66, 9.3$ $(0.02, 4.4)$	0.46	t $0.66, 9.2$ $(0.02, 4.4)$	0.46	N 0.66 (0.02)	0.46
C_{42}	S t	$0.57, 17.4$ $(0.02, 4.0)$	0.39	t $0.57, 17.4$ $(0.02, 4.0)$	0.39	N 0.57 (0.02)	0.39
	J t	$0.56, 14.0$ $(0.02, 4.4)$	0.38	t $0.56, 15.6$ $(0.02, 4.2)$	0.38	N 0.57 (0.02)	0.39
$C_{43\|2}$	S t	$-0.29, 8.9$ $(0.03, 3.4)$	-0.19	t $-0.30, 9.8$ $(0.03, 4.3)$	-0.20	N -0.28 (0.04)	-0.18
	J t	$-0.29, 8.7$ $(0.03, 3.6)$	-0.19	t $-0.30, 9.5$ $(0.03, 4.3)$	-0.20	N -0.28 (0.04)	-0.18

Table 5.2: Sequential (S) and joint (J) ML estimates, standard errors (parentheses), and estimates of Kendall's τ for the Gaussian, non-Gaussian, and semiparametric non-Gaussian PCBNs applied to the log-return data. Copulas include the Gaussian (N) and Student's t (t) pair-copula families.

PCBN		LL	# Parameters	AIC
n$\mathcal{G}$	S	509.0	8	-1002.0
	J	509.2	8	-1002.4
sn$\mathcal{G}$	S	507.8	8	-999.6
	J	507.9	8	-999.8
$\mathcal{G}$	S	489.7	4	-971.5
	J	489.8	4	-971.5

Table 5.3: Maximised log-likelihoods, numbers of parameters, and AIC values for the Gaussian ($\mathcal{G}$), non-Gaussian (n$\mathcal{G}$), and semiparametric non-Gaussian (sn$\mathcal{G}$) PCBNs applied to the log-return data. Sequential (S) and joint (J) ML estimates of the corresponding parameters are given in Table 5.2.

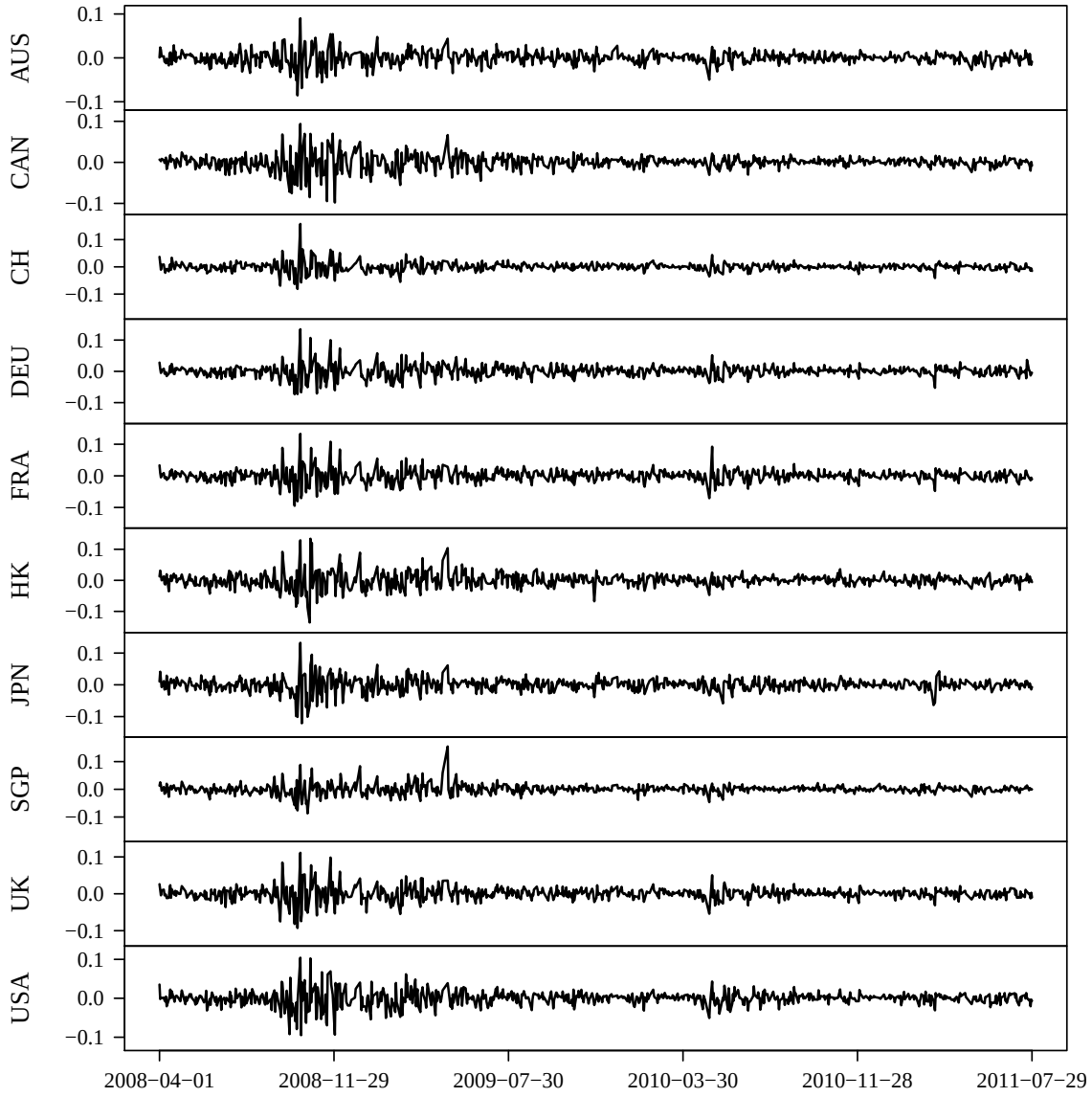


Figure 5.3: Daily log-returns of the Australian All Ordinaries (AUS), the Canadian S&P/TSX Composite Index (CAN), the Swiss Market Index (CH), the German DAX (DEU), the French CAC 40 (FRA), the Hong Kong Hang Seng Index (HK), the Japanese Nikkei 225 (JPN), the Singapore Straits Times Index (SGP), the UK FTSE 100 (UK), & the US S&P 500 (USA).

	μ [$\times 10^3$]	a	ω [$\times 10^5$]	α	β	ν	γ
AUS	0.26 (0.41)	0.01 (0.04)	0.18 (3.77)	0.08 (0.02)	0.91 (0.02)	10.33 (3.46)	0.90 (0.05)
CAN	0.50 (0.38)	−0.01 (0.04)	0.17 (3.83)	0.09 (0.02)	0.90 (0.02)	10.74 (3.73)	0.80 (0.05)
CH	0.08 (0.38)	0.02 (0.04)	0.38 (4.01)	0.12 (0.03)	0.86 (0.02)	7.64 (2.04)	0.92 (0.05)
DEU	0.66 (0.49)	−0.03 (0.04)	0.31 (4.03)	0.08 (0.02)	0.91 (0.02)	7.77 (2.39)	0.93 (0.05)
FRA	0.23 (0.54)	−0.02 (0.04)	0.59 (4.24)	0.09 (0.02)	0.89 (0.02)	8.29 (2.44)	0.94 (0.05)
HK	0.36 (0.55)	−0.02 (0.04)	0.23 (4.07)	0.07 (0.02)	0.93 (0.01)	8.23 (2.25)	0.97 (0.05)
JPN	0.27 (0.53)	−0.06 (0.04)	0.85 (4.39)	0.12 (0.03)	0.85 (0.02)	15.33 (7.85)	0.87 (0.05)
SGP	0.57 (0.40)	−0.01 (0.04)	0.24 (3.98)	0.09 (0.02)	0.90 (0.02)	4.92 (0.89)	1.03 (0.05)
UK	0.56 (0.44)	−0.01 (0.04)	0.37 (4.02)	0.09 (0.02)	0.89 (0.02)	8.42 (2.60)	0.92 (0.05)
USA	0.76 (0.43)	−0.07 (0.04)	0.23 (4.03)	0.10 (0.02)	0.89 (0.02)	7.22 (2.15)	0.84 (0.04)

Table 5.4: ML estimates and standard errors (in parentheses) of AR(1)-GARCH(1,1) parameters for the ten time series of daily log-returns.

are assumed to follow a skewed Student's t distribution with ν_i degrees of freedom and skewness parameter γ_i, see McNeil et al. (2005, Section 3.2). The corresponding cdf will be denoted by $\mathrm{t}_{\nu_i,\gamma_i}$. ML parameter estimates and corresponding standard errors derived from numerical evaluation of the Hessian of the AR(1)-GARCH(1,1) parameters are given in Table 5.4. We assessed model fit using the following statistical tests: the Ljung-Box test (Ljung and Box, 1978) with null hypothesis that there is no autocorrelation left in the residuals and squared residuals, the Langrange-multiplier ARCH test (Engle, 1982) with null hypothesis that the residuals exhibit no conditional heteroskedasticity, and the Kolmogorov-Smirnov test (Conover, 1999, Section 6.1) with null hypothesis that the residuals follow a skewed Student's t distribution. None of these null hypotheses could be rejected at the 5% significance level. We then transformed the standardised residuals to uniformly distributed observations $u_{t,i} := \mathrm{t}_{\nu_i,\gamma_i}\left(\sqrt{\frac{\nu_i}{\nu_i-2} + \frac{2\,\nu_i^2\,\gamma_i^2}{(\nu_i-2)^2\,(\nu_i-4)}}\, z_{t,i}\right)$, before modelling the joint dependence structure of the ten time series of log-returns by a PCBN.

5.2.2 Estimating the conditional independence structure

We estimated the conditional independence structure of the ten time series of log-returns by applying the PC algorithm with either of the ten conditional independence tests COR, C-GR, C-H, C-K, D-GR, D-H, D-K, R-GR, R-H, and R-K described in

Section 4.1.3 (with notation as in Section 4.2) to the transformed observations $u_{t,i}$. All tests were performed at the 5% level. As a result, we obtained three different essential graphs $\mathcal{D}^e_{\mathrm{COR}}$, $\mathcal{D}^e_{\mathrm{GR,H}}$, and $\mathcal{D}^e_{\mathrm{K}}$, of which the first was returned by the PC algorithm with COR, the second was returned by the PC algorithm with either of C-GR, C-H, D-GR, D-H, R-GR, and R-H, and the third was returned by the PC algorithm with either of C-K, D-K, and R-K, respectively. Obviously, a restriction of the class of R-vines to C- or D-vines had not influence on the resulting essential graph. We then oriented undirected edges in the obtained essential graphs, as described in Section 1.2, in order to obtain DAGs $\mathcal{D}_{\mathrm{COR}}$, $\mathcal{D}_{\mathrm{GR,H}}$, and $\mathcal{D}_{\mathrm{K}}$ from the Markov-equivalence classes represented by $\mathcal{D}^e_{\mathrm{COR}}$, $\mathcal{D}^e_{\mathrm{GR,H}}$, and $\mathcal{D}^e_{\mathrm{K}}$, respectively. More precisely, $\mathcal{D}^e_{\mathrm{COR}}$ contained the two undirected edges AUS $-$ HK and CH $-$ DEU, which we replaced by AUS $\to$ HK and CH $\to$ DEU, respectively, based on the heuristic rule that $\mathcal{D}^e_{\mathrm{GR,H}}$ and $\mathcal{D}^e_{\mathrm{K}}$ already contained AUS $\to$ HK and CH $\to$ DEU. Similarly, we oriented AUS $-$ JPN into AUS $\leftarrow$ JPN in $\mathcal{D}_{\mathrm{GR,H}}$ and $\mathcal{D}_{\mathrm{K}}$ since $\mathcal{D}^e_{\mathrm{COR}}$ already contained AUS $\leftarrow$ JPN. The DAGs $\mathcal{D}_{\mathrm{COR}}$, $\mathcal{D}_{\mathrm{GR,H}}$, and $\mathcal{D}_{\mathrm{K}}$ are given in Figure 5.4.

In all three DAGs in Figure 5.4, the Asian-Pacific indices AUS, HK, JPN, and SGP are mutually adjacent, and so are the two North American indices CAN and USA. The same holds true for the European indices CH, DEU, FRA, and UK in DAG $\mathcal{D}_{\mathrm{COR}}$, while DEU and UK are non-adjacent in $\mathcal{D}_{\mathrm{GR,H}}$ and $\mathcal{D}_{\mathrm{K}}$. A probability measure satisfying the Markov properties represented by either $\mathcal{D}_{\mathrm{GR,H}}$ or $\mathcal{D}_{\mathrm{K}}$, respectively, observes the conditional independence restriction DEU $\perp\!\!\!\perp$ UK $\mid \{\mathrm{CH}, \mathrm{FRA}\}$. All further conditional independence restrictions represented by the DAGs in Figure 5.4 involve indices in at least two of the above given regions Asia-Pacific, Europe, and North America. We hence observe a strong geographical clustering of dependences. Moreover, all three DAGs in Figure 5.4 represent the conditional independence restriction $\{\mathrm{AUS}, \mathrm{HK}, \mathrm{JPN}, \mathrm{SGP}\} \perp\!\!\!\perp \{\mathrm{CAN}, \mathrm{USA}\} \mid \{\mathrm{CH}, \mathrm{DEU}, \mathrm{FRA}, \mathrm{UK}\}$, that is, Asia-Pacific $\perp\!\!\!\perp$ North America $\mid$ Europe. Note that Markov properties alone are not sufficient for deriving causal relations within the analysed data (see, for instance, the undirected edges in an essential graph), but they can be used as a starting point for further research in that direction.

A well-ordering for $\mathcal{D}_{\text{COR}}$ is given by $1 \mapsto$ CAN, $2 \mapsto$ CH, $3 \mapsto$ DEU, $4 \mapsto$ UK, $5 \mapsto$ FRA, $6 \mapsto$ USA, $7 \mapsto$ JPN, $8 \mapsto$ SGP, $9 \mapsto$ AUS, $10 \mapsto$ HK. Similarly, we obtain a well-ordering for $\mathcal{D}_{\text{GR,H}}$ and $\mathcal{D}_{\text{K}}$, respectively, by mapping $1 \mapsto$ CAN, $2 \mapsto$ CH, $3 \mapsto$ UK, $4 \mapsto$ FRA, $5 \mapsto$ DEU, $6 \mapsto$ USA, $7 \mapsto$ JPN, $8 \mapsto$ AUS, $9 \mapsto$ SGP, $10 \mapsto$ HK. We determined parent orderings for the three DAGs in Figure 5.4 in two steps. First, we applied the greedy-type procedure with Kendall's τ edge weights described in Section 4.3, and second, we permuted some of the orderings obtained in step one to reduce the number of integrals in the corresponding pair-copula decompositions and thus the computational complexity. More precisely, we changed JPN $<_{\text{AUS}}$ SGP and CAN $<_{\text{USA}}$ DEU $<_{\text{USA}}$ FRA in DAG $\mathcal{D}_{\text{COR}}$ into SGP $<_{\text{AUS}}$ JPN and DEU $<_{\text{USA}}$ FRA $<_{\text{USA}}$ CAN, respectively, and CAN $<_{\text{USA}}$ DEU $<_{\text{USA}}$ FRA in DAG $\mathcal{D}_{\text{K}}$ into DEU $<_{\text{USA}}$ FRA $<_{\text{USA}}$ CAN. The resulting parent orderings for $\mathcal{D}_{\text{COR}}$, $\mathcal{D}_{\text{GR,H}}$, and $\mathcal{D}_{\text{K}}$, respectively, are displayed in Figure 5.4.

5.2.3 Pair-copula selection and ML estimation

Having fixed the parent orderings for the three PCBNs corresponding to $\mathcal{D}_{\text{COR}}$, $\mathcal{D}_{\text{GR,H}}$, and $\mathcal{D}_{\text{K}}$, respectively, we next selected parametric pair-copula families using the AIC as a selection criterion. We considered the Clayton, Frank, Gaussian, Gumbel, and Student's t copula families as well as reflected versions of the Clayton and Gumbel copula families in order to account for negative correlations. We then computed sequential ML estimates of the parameters of the so specified PCBNs. Selected pair-copula families, corresponding sequential ML estimates, bootstrapped standard errors, and estimates of Kendall's τ are given in Table 5.5. The respective maximised log-likelihoods and AIC values are summarised in Table 5.6. Moreover, we compared model fit to the respective Gaussian PCBNs comprising only Gaussian pair copulas. Corresponding ML estimates, standard errors, and estimates of Kendall's τ are again found in Table 5.5, while maximised log-likelihoods and AIC values are given in Table 5.6.

According to the AIC, the best fit was obtained by the non-Gaussian PCBN with DAG $\mathcal{D}_{\text{GR,H}}$, followed by the non-Gaussian PCBNs associated to $\mathcal{D}_{\text{K}}$ and $\mathcal{D}_{\text{COR}}$,

DAG			$\mathcal{D}_{\text{COR}}$			$\mathcal{D}_{\text{GR,H}}$			$\mathcal{D}_{\text{K}}$	
			Parameters	$\widehat{\tau}$		Parameters	$\widehat{\tau}$		Parameters	$\widehat{\tau}$
JPN → AUS	n$\mathcal{G}$	t	0.56, 10.3 (0.03, 4.1)	0.38	t	0.73, 8.5 (0.02, 3.4)	0.52	t	0.73, 8.5 (0.02, 3.4)	0.52
	$\mathcal{G}$	N	0.55 (0.03)	0.37	N	0.72 (0.02)	0.52	N	0.72 (0.02)	0.52
SGP → AUS	n$\mathcal{G}$	t	0.64, 9.4 (0.02, 4.1)	0.44						
	$\mathcal{G}$	N	0.64 (0.02)	0.44						
CH → DEU	n$\mathcal{G}$	t	0.83, 5.4 (0.01, 1.8)	0.63	F	0.98 (0.23)	0.11	F	0.98 (0.23)	0.11
	$\mathcal{G}$	N	0.82 (0.01)	0.62	N	0.14 (0.04)	0.09	N	0.14 (0.04)	0.09
FRA → DEU	n$\mathcal{G}$				t	0.94, 3.8 (0.01, 1.2)	0.78	t	0.94, 3.8 (0.01, 1.2)	0.78
	$\mathcal{G}$				N	0.93 (0.01)	0.76	N	0.93 (0.01)	0.76
CH → FRA	n$\mathcal{G}$	F	1.55 (0.24)	0.17	t	0.44, 9.0 (0.03, 4.4)	0.29	t	0.44, 9.0 (0.03, 4.3)	0.29
	$\mathcal{G}$	N	0.28 (0.04)	0.18	N	0.44 (0.03)	0.29	N	0.44 (0.03)	0.29
DEU → FRA	n$\mathcal{G}$	t	0.94, 3.8 (0.01, 1.1)	0.78						
	$\mathcal{G}$	N	0.93 (0.01)	0.76						
UK → FRA	n$\mathcal{G}$	t	0.57, 6.9 (0.03, 3.3)	0.38	t	0.92, 7.3 (0.01, 2.8)	0.74	t	0.92, 7.3 (0.01, 2.8)	0.74
	$\mathcal{G}$	N	0.59 (0.04)	0.40	N	0.92 (0.01)	0.74	N	0.92 (0.01)	0.74
AUS → HK	n$\mathcal{G}$	t	0.35, 13.4 (0.03, 4.8)	0.22	t	0.35, 13.4 (0.03, 4.8)	0.22	t	0.35, 13.4 (0.03, 4.8)	0.22
	$\mathcal{G}$	N	0.36 (0.04)	0.24	N	0.36 (0.04)	0.24	N	0.36 (0.04)	0.24
JPN → HK	n$\mathcal{G}$	t	0.20, 20.0 (0.04, 2.8)	0.13	N	0.20 (0.04)	0.13	N	0.20 (0.04)	0.13
	$\mathcal{G}$	N	0.20 (0.04)	0.13	N	0.20 (0.04)	0.13	N	0.20 (0.04)	0.13
SGP → HK	n$\mathcal{G}$	t	0.78, 5.7 (0.01, 1.9)	0.57	t	0.78, 5.7 (0.01, 2.0)	0.57	t	0.78, 5.7 (0.01, 2.0)	0.57
	$\mathcal{G}$	N	0.78 (0.02)	0.57	N	0.78 (0.02)	0.57	N	0.78 (0.02)	0.57
AUS → SGP	n$\mathcal{G}$				t	0.64, 9.4 (0.02, 4.0)	0.44	t	0.64, 9.4 (0.02, 4.0)	0.44
	$\mathcal{G}$				N	0.64 (0.02)	0.44	N	0.64 (0.02)	0.44
JPN → SGP	n$\mathcal{G}$	t	0.60, 11.6 (0.02, 4.5)	0.41	N	0.27 (0.03)	0.17	N	0.27 (0.03)	0.17
	$\mathcal{G}$	N	0.60 (0.02)	0.41	N	0.26 (0.03)	0.17	N	0.26 (0.04)	0.17
UK → SGP	n$\mathcal{G}$	N	0.34 (0.03)	0.22	N	0.27 (0.03)	0.18	N	0.27 (0.03)	0.18
	$\mathcal{G}$	N	0.34 (0.03)	0.22	N	0.27 (0.03)	0.18	N	0.27 (0.03)	0.18
CAN → UK	n$\mathcal{G}$	SG	1.13 (0.03)	0.12	N	0.31 (0.03)	0.20	N	0.31 (0.03)	0.20
	$\mathcal{G}$	N	0.21 (0.04)	0.13	N	0.31 (0.03)	0.20	N	0.31 (0.03)	0.20
CH → UK	n$\mathcal{G}$	t	0.36, 9.6 (0.03, 4.5)	0.23	t	0.83, 8.6 (0.01, 3.7)	0.62	t	0.83, 8.6 (0.01, 3.7)	0.62
	$\mathcal{G}$	N	0.39 (0.04)	0.25	N	0.83 (0.01)	0.62	N	0.83 (0.01)	0.62
DEU → UK	n$\mathcal{G}$	t	0.88, 7.3 (0.01, 3.0)	0.69						
	$\mathcal{G}$	N	0.88 (0.01)	0.68						
CAN → USA	n$\mathcal{G}$	N	0.48 (0.03)	0.32	t	0.75, 6.7 (0.02, 3.0)	0.54	N	0.52 (0.03)	0.35
	$\mathcal{G}$	N	0.48 (0.03)	0.32	N	0.75 (0.01)	0.54	N	0.52 (0.03)	0.35
DEU → USA	n$\mathcal{G}$	t	0.71, 6.8 (0.02, 3.0)	0.51	t	0.47, 12.7 (0.03, 4.9)	0.31	t	0.71, 6.8 (0.02, 3.0)	0.51
	$\mathcal{G}$	N	0.71 (0.02)	0.50	N	0.47 (0.03)	0.31	N	0.71 (0.02)	0.50
FRA → USA	n$\mathcal{G}$	t	0.19, 9.2 (0.04, 4.5)	0.12				t	0.19, 9.2 (0.04, 4.4)	0.12
	$\mathcal{G}$	N	0.21 (0.05)	0.14				N	0.21 (0.05)	0.14

Table 5.5: Selected pair-copula families, sequential ML estimates, standard errors (parentheses), and estimates of Kendall's τ for the Gaussian ($\mathcal{G}$) and non-Gaussian (n$\mathcal{G}$) PCBNs corresponding to the DAGs in Figure 5.4. Copulas include the Frank (F), Gaussian (N), Survival Gumbel (SG), and Student's t (t) pair-copula families.

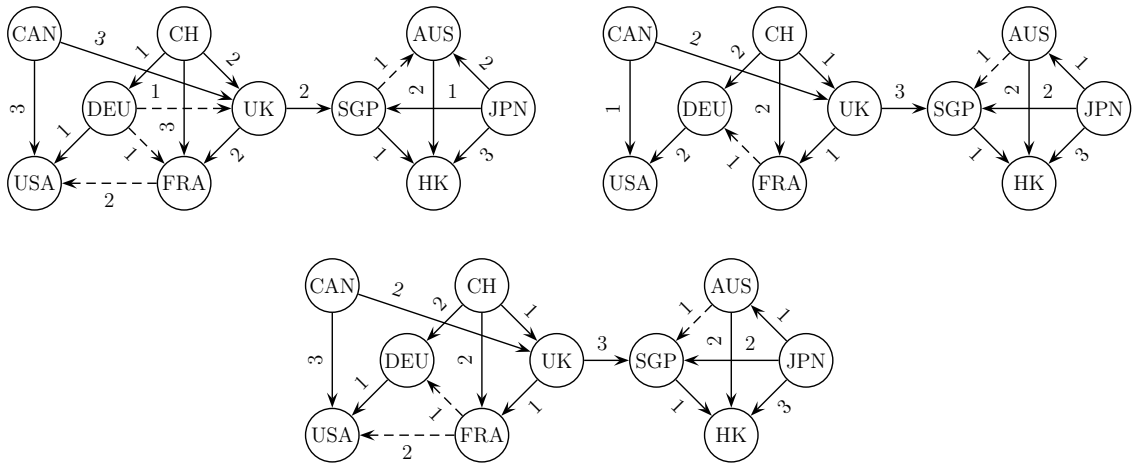


Figure 5.4: DAGs $\mathcal{D}_{\text{COR}}$ (top left), $\mathcal{D}_{\text{GR,H}}$ (top right), $\mathcal{D}_{\text{K}}$ (bottom) returned by the PC algorithm with different conditional independence tests when estimating the Markov structure of the ten time series AUS, CAN, CH, DEU, FRA, HK, JPN, SGP, UK, USA of daily log-returns. Solid edges appear in all three DAGs. Edge labels indicate parent orderings, that is, for instance, DEU $<_{\text{USA}}$ FRA $<_{\text{USA}}$ CAN in $\mathcal{D}_{\text{COR}}$.

DAG		LL	# Parameters	AIC
$\mathcal{D}_{\text{COR}}$	n$\mathcal{G}$	3397.0	30	−6734.0
	$\mathcal{G}$	3264.6	17	−6495.3
$\mathcal{D}_{\text{GR,H}}$	n$\mathcal{G}$	3412.8	25	−6775.6
	$\mathcal{G}$	3285.5	15	−6540.9
$\mathcal{D}_{\text{K}}$	n$\mathcal{G}$	3401.9	26	−6751.8
	$\mathcal{G}$	3264.1	16	−6496.3

Table 5.6: Maximised log-likelihoods, numbers of parameters, and AIC values for the Gaussian ($\mathcal{G}$) and non-Gaussian (n$\mathcal{G}$) PCBNs corresponding to the DAGs in Figure 5.4. ML estimates of the parameters are given in Table 5.5.

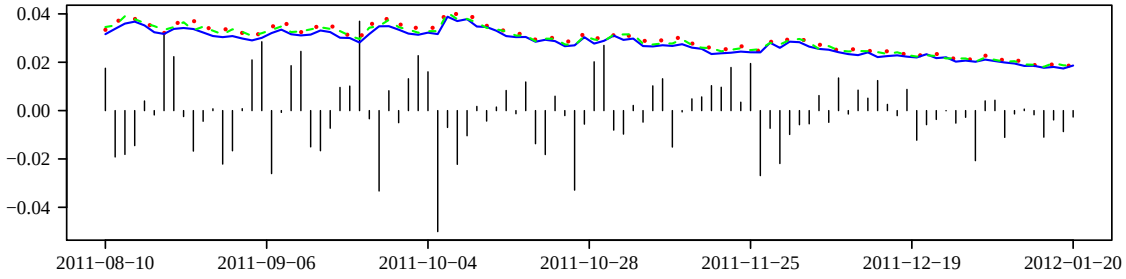


Figure 5.5: Predicted one-day 99% VaR based on 100,000 samples from either of the non-Gaussian PCBNs with DAGs $\mathcal{D}_{\text{COR}}$ (dotted), $\mathcal{D}_{\text{GR,H}}$ (solid), $\mathcal{D}_{\text{K}}$ (dashed). Bars indicate true portfolio losses.

respectively. Applying the Vuong test with AIC correction (Vuong, 1989) to the non-Gaussian PCBNs at the 5% level, we cannot reject the null hypothesis that all three models are equally close to the true model. A similar statement holds for the Gaussian PCBNs. However, using the Vuong test for model selection between a Gaussian and a non-Gaussian PCBN will always decide in favor of the non-Gaussian model, which again shows the latter model's superiority. This is, of course, to be expected since financial returns often exhibit heavy-tailed dependence, which is validated here by the low estimates of the degrees-of-freedom parameters of the t copulas.

5.2.4 Value-at-Risk prediction

We finally examined the out-of-sample performance of the non-Gaussian PCBNs by predicting Value-at-Risk (VaR) (McNeil et al., 2005, Section 2.2) for the period from 10 August 2011 to 20 January 2012 ($N = 100$ observations). To this end, we considered an equally weighted portfolio of above stock indices, and performed the following steps at each time point t. First, we set the portfolio value at time $t-1$ to 1 and computed the true portfolio loss $L_t := 1 - \frac{1}{10}\sum_{i=1}^{10} e^{r_{t,i}}$. Second, we drew 100,000 samples from either of the three non-Gaussian PCBNs under investigation. Last, we used the freshly drawn samples to predict the one-day 99% VaR and compared the obtained values to L_t. The results are displayed in Figure 5.5.

The VaR predictions of all three models are almost indistinguishable and are each

only once exceeded by the true portfolio losses. Hence, for either model, the null hypotheses of correct unconditional (Kupiec, 1995) and conditional (Christoffersen, 1998) coverage, respectively, cannot be rejected at the 5% significance level.

5.3 Reducing computational complexity

Depending on the underlying DAG and the corresponding parent orderings, the evaluation of the log-likelihood of a PCBN may involve high-dimensional numerical integration and hence considerable computational effort. In Section 4.3, we presented a greedy procedure for selecting the parent orderings of the vertices of the underlying DAG, which is based on the idea of modelling strongest dependences in the pair copulas with fewest conditioning variables. In Section 5.2.2, we introduced an additional selection step, in which some of the parent sets were rearranged in order to reduce the number of integrals in the corresponding likelihood decompositions. It would be desirable to have theoretical results on the relationship between parent orderings and the number and complexity of integrals in the future. Moreover, in Section 5.1.3, we suggested to replace some or all of the integrals by non-parametric kernel conditional cdf estimators. Another way of reducing computational complexity is to consider sequential instead of joint ML estimates.

A H-functions of selected copula families

The h-functions of the Clayton, Frank, Gumbel, Gaussian, and Student's t pair-copula families discussed in Section 2.1 exhibit closed form expressions.

Clayton copula

$$h_{\underline{1}2}(\boldsymbol{u};\theta) = \left((u_1^{-\theta}-1)\,u_2^{\theta}+1\right)^{-1-\frac{1}{\theta}}, \quad \boldsymbol{u}=(u_1,u_2)\in[0,1]^2, \quad \theta\in(0,\infty).$$

Frank copula

$$h_{\underline{1}2}(\boldsymbol{u};\theta) = \frac{(e^{-\theta\,u_1}-1)\,e^{-\theta\,u_2}}{(e^{-\theta\,u_1}-1)\,(e^{-\theta\,u_2}-1)+e^{-\theta}-1}, \quad \theta\in\mathbb{R}\setminus\{0\}.$$

Gumbel copula

$$h_{\underline{1}2}(\boldsymbol{u};\theta) = \frac{C(\boldsymbol{u};\theta)}{u_2}\left(1+\left(\frac{\log u_1}{\log u_2}\right)^{\theta}\right)^{-1+\frac{1}{\theta}}, \quad \theta\in[1,\infty).$$

Gaussian copula

$$h_{\underline{1}2}(\boldsymbol{u};\rho) = \Phi\left(\frac{\Phi^{-1}(u_1)-\rho\;\Phi^{-1}(u_2)}{\sqrt{1-\rho^2}}\right), \quad \rho\in(-1,1).$$

Student's t copula

$$h_{\underline{1}2}(\boldsymbol{u};\rho,\nu) = \mathrm{t}_{\nu+1}\left(\frac{\sqrt{\nu+1}\left(\mathrm{t}_{\nu}^{-1}(u_1)-\rho\ \mathrm{t}_{\nu}^{-1}(u_2)\right)}{\sqrt{(1-\rho^2)\left(\nu+\left[\mathrm{t}_{\nu}^{-1}(u_2)\right]^2\right)}}\right), \quad \rho\in(-1,1), \nu\in(1,\infty).$$

B Example of a PCC for a D-Markovian probability measure

In Examples 3.2 to 3.4, we considered a $\mathcal{D}$-Markovian probability measure on $\mathbb{R}^7$ associated to the DAG $\mathcal{D}$ in Figure 3.2. We decomposed the pdf according to

$$\begin{aligned} f = f_1 \cdots f_7 \, & c_{21}(F_2, F_1) \, c_{31}(F_3, F_1) \, c_{42}(F_4, F_2) \, c_{41|2}(F_{4|2}, F_{1|2}) \, c_{54}(F_5, F_4) \, c_{53|4}(F_{5|4}, F_{3|4}) \\ & \cdot c_{65}(F_6, F_5) \, c_{64|5}(F_{6|5}, F_{4|5}) \, c_{63|54}(F_{6|54}, F_{3|54}) \, c_{62|543}(F_{6|543}, F_{2|543}) \, c_{75}(F_7, F_5) \\ & \cdot c_{76|5}(F_{7|5}, F_{6|5}) \, c_{73|56}(F_{7|56}, F_{3|56}). \end{aligned}$$

We then used Algorithm 2 to derive a pair-copula decomposition for the conditional cdf $F_{3|56}$. Applying Algorithm 2 also to the other conditional cdfs yields

$$\begin{aligned} F_{1|2} &= h_{2\underline{1}}(F_2, F_1), \\ F_{4|2} &= h_{\underline{4}2}(F_4, F_2), \\ F_{3|4} &= \int_{\mathbb{R}^2} c_{41|2}(F_{4|2}, F_{1|2}) \, c_{42}(F_4, F_2) \, h_{\underline{3}1}(F_3, F_1) \, f_2 \, c_{21}(F_2, F_1) \, f_1 \, \mathrm{d}\boldsymbol{x}_{12}, \\ F_{5|4} &= h_{\underline{5}4}(F_5, F_4), \\ F_{4|5} &= h_{5\underline{4}}(F_5, F_4), \\ F_{6|5} &= h_{\underline{6}5}(F_6, F_5), \\ F_{7|5} &= h_{\underline{7}5}(F_7, F_5), \\ F_{3|54} &= h_{5\underline{3}|4}(F_{5|4}, F_{3|4}), \end{aligned}$$

$$
\begin{aligned}
F_{6|54} &= h_{\underline{6}4|5}(F_{6|5}, F_{4|5}),\\
F_{3|56} &= \int_{\mathbb{R}} h_{6\underline{3}|54}(F_{6|54}, F_{3|54})\, c_{64|5}(F_{6|5}, F_{4|5})\, c_{54}(F_5, F_4)\, f_4 \,\mathrm{d}x_4,\\
F_{7|56} &= h_{\underline{7}6|5}(F_{7|5}, F_{6|5}),\\
F_{2|543} &= \frac{\int_{-\infty}^{\cdot} \int_{\mathbb{R}} c_{41|2}(F_{4|2}, F_{1|2})\, c_{42}(F_4, F_2)\, c_{31}(F_3, F_1)\, f_2\, c_{21}(F_2, F_1)\, f_1 \,\mathrm{d}x_1 \,\mathrm{d}x_2}{\int_{\mathbb{R}^2} c_{41|2}(F_{4|2}, F_{1|2})\, c_{42}(F_4, F_2)\, c_{31}(F_3, F_1)\, f_2\, c_{21}(F_2, F_1)\, f_1 \,\mathrm{d}\boldsymbol{x}_{12}},\\
F_{6|543} &= h_{\underline{6}3|54}(F_{6|54}, F_{3|54}).
\end{aligned}
$$

Bibliography

K. Aas and D. Berg. Models for construction of multivariate dependence—a comparison study. *The European Journal of Finance*, 15(7–8):639–659, 2009.

K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44:182–198, 2009.

E. F. Acar, C. Genest, and J. Nešlehová. Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110:74–90, 2012.

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Chichester, third edition, 2003.

S. A. Andersson, D. Madigan, and M. D. Perlman. A characterization of Markov equivalence classes for acyclic digraphs. *The Annals of Statistics*, 25(2):505–541, 1997.

D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook. *The Traveling Salesman Problem: A Computational Study*. Princeton University Press, Princeton, New Jersey, 2007.

A. Bauer and C. Czado. Pair-copula Bayesian networks. Submitted for publication, arXiv:1211.5620 [stat.ME], 2012.

A. Bauer, C. Czado, and T. Klein. Pair-copula constructions for non-Gaussian DAG models. *The Canadian Journal of Statistics*, 40(1):86–109, 2012.

T. Bedford and R. M. Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32:245–268, 2001.

T. Bedford and R. M. Cooke. Vines—a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068, 2002.

W. P. Bergsma. Nonparametric testing of conditional independence by means of the partial copula. arXiv:1101.4607 [math.ST], 2011.

J. R. Blum, J. Kiefer, and M. Rosenblatt. Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, 32(2):485–498, 1961.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

E. C. Brechmann, C. Czado, and K. Aas. Truncated regular vines in high dimensions with application to financial data. *The Canadian Journal of Statistics*, 40(1):68–85, 2012.

U. Cherubini, E. Luciano, and W. Vecchiato. *Copula Methods in Finance*. John Wiley & Sons, Chichester, 2004.

U. Cherubini, F. Gobbi, S. Mulinacci, and S. Romagnoli. *Dynamic Copula Methods in Finance*. John Wiley & Sons, Chichester, 2011.

L. Chollete, A. Heinen, and A. Valdesogo. Modeling international financial returns with a multivariate regime-switching copula. *Journal of Financial Econometrics*, 7(4):437–480, 2009.

P. F. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39(4):841–862, 1998.

W. J. Conover. *Practical Nonparametric Statistics*. John Wiley & Sons, New York, third edition, 1999.

R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*. Springer, New York, second edition, 2003.

D. R. Cox and N. Wermuth. *Multivariate Dependencies: Models, Analysis and Interpretation*. Chapman & Hall, London, 1996.

C. Czado, U. Schepsmeier, and A. Min. Maximum likelihood estimation of mixed C-vines with application to exchange rates. *Statistical Modelling*, 12(3):229–255, 2012.

J. Dißmann, E. C. Brechmann, C. Czado, and D. Kurowicka. Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics and Data Analysis*, 59:52–69, 2013.

D. Dor and M. Tarsi. A simple algorithm to construct a consistent extension of a partially oriented graph. Technical report R-185, Cognitive Systems Laboratory, UCLA, 1992.

G. Elidan. Copula Bayesian networks. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta (Eds.), *Advances in Neural Information Processing Systems*, volume 23, pages 559–567. NIPS Foundation, La Jolla, California, 2010.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

M. Fischer, C. Köck, S. Schlüter, and F. Weigert. An empirical analysis of multivariate copula models. *Quantitative Finance*, 9(7):839–854, 2009.

C. Genest and B. Rémillard. Tests of independence and randomness based on the empirical copula process. *Test*, 13(2):335–369, 2004.

C. Genest, K. Ghoudi, and L.-P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82 (3):543–552, 1995.

L. Gruber, C. Czado, and J. Stöber. Bayesian model selection for regular vine copulas using reversible jump MCMC. Submitted for publication, 2013.

A. M. Hanea. *Algorithms for non-parametric Bayesian belief nets.* Ph.D. thesis, Technische Universiteit Delft, 2008.

A. M. Hanea and D. Kurowicka. Mixed non-parametric continuous and discrete Bayesian belief nets. In T. Bedford, J. Quigley, L. Walls, B. Alkali, A. Daneshkhah, and G. Hardman (Eds.), *Advances in Mathematical Modeling for Reliability*, pages 9–16. IOS Press, Amsterdam, 2008.

A. M. Hanea, D. Kurowicka, and R. M. Cooke. Hybrid method for quantifying and analyzing Bayesian belief nets. *Quality and Reliability Engineering International*, 22:709–729, 2006.

N. Harris and M. Drton. PC algorithm for Gaussian copula graphical models. arXiv:1207.0242 [math.ST], 2012.

I. Hobæk Haff. Comparing estimators for pair-copula constructions. *Journal of Multivariate Analysis*, 110:91–105, 2012.

I. Hobæk Haff. Parameter estimation for pair-copula constructions. *Bernoulli*, 2013. To appear.

I. Hobæk Haff, K. Aas, and A. Frigessi. On the simplified pair-copula construction—Simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5):1296–1310, 2010.

W. Hoeffding. A non-parametric test of independence. *The Annals of Mathematical Statistics*, 19(4):546–557, 1948.

M. Hollander and D. A. Wolfe. *Nonparametric Statistical Methods.* John Wiley & Sons, Chichester, second edition, 1999.

K. Ignatieva and E. Platen. Modelling co-movements and tail dependency in the international stock market via copulae. *Asia-Pacific Financial Markets*, 17(3): 261–302, 2010.

H. Joe. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In L. Rüschendorf, B. Schweizer, and M. D. Taylor (Eds.), *Distributions with Fixed Marginals and Related Topics*, volume 28 of *Lecture Notes—Monograph Series*, pages 120–141. Institute of Mathematical Statistics, Hayward, California, 1996.

H. Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall, London, 1997.

H. Joe and J. J. Xu. The estimation method of inference functions for margins for multivariate models. Technical report 166, Department of Statistics, University of British Columbia, 1996.

M. Kalisch and P. Bühlmann. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research*, 8:613–636, 2007.

M. Kalisch, M. Mächler, and D. Colombo. *Estimation of CPDAG/PAG and causal inference using the IDA algorithm*, 2012. Manual for the R package `pcalg`.

G. Kim, M. J. Silvapulle, and P. Silvapulle. Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics and Data Analysis*, 51(6):2836–2850, 2007.

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge, Massachusetts, 2009.

J. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50, 1956.

P. H. Kupiec. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2):73–84, 1995.

D. Kurowicka. Optimal Truncation of Vines. In D. Kurowicka and H. Joe (Eds.), *Dependence Modeling: Vine Copula Handbook*, pages 233–248. World Scientific, Singapore, 2011.

D. Kurowicka and R. Cooke. *Uncertainty Analysis with High Dimensional Dependence Modelling*. John Wiley & Sons, Chichester, 2006.

D. Kurowicka and R. M. Cooke. Distribution-free continuous Bayesian belief nets. In A. Wilson, N. Limnios, S. Keller-McNulty, and Y. Armijo (Eds.), *Modern Statistical and Mathematical Methods in Reliability*, volume 10 of *Series on Quality, Reliability and Engineering Statistics*, pages 309–323. World Scientific, Singapore, 2005.

D. Kurowicka and H. Joe (Eds.). *Dependence Modeling: Vine Copula Handbook*. World Scientific, Singapore, 2011.

S. L. Lauritzen. *Graphical Models*. Oxford University Press, Oxford, 1996.

Q. Li and J. S. Racine. *Nonparamatric Econometrics: Theory and Practice*. Princeton University Press, Princeton, New Jersey, 2007.

G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, New Jersey, 2005.

C. Meek. Causal inference and causal explanation with background knowledge. In P. Besnard and S. Hanks (Eds.), *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 403–410. Morgan Kaufmann, San Francisco, California, 1995.

A. Min and C. Czado. Bayesian model selection for D-vine pair-copula constructions. *The Canadian Journal of Statistics*, 39(2):239–258, 2011.

B. R. Moole and M. Valtorta. Sequential and parallel algorithms for causal explanation with background knowledge. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12(Supplement 2):101–122, 2004.

R. E. Neapolitan. *Learning Bayesian Networks*. Prentice Hall, Upper Saddle River, New Jersey, 2003.

R. B. Nelsen. *An Introduction to Copulas*. Springer, New York, second edition, 2006.

A. J. Patton. Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556, 2006.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, second edition, 2009.

O. Pourret, P. Naïm, and B. Marcot (Eds.). *Bayesian Networks: A Practical Guide to Applications*. John Wiley & Sons, Chichester, 2008.

R. C. Prim. Shortest connection networks and some generalizations. *Bell System Technical Journal*, 36(6):1389–1401, 1957.

R. W. Robinson. Counting labeled acyclic digraphs. In F. Harary (Ed.), *New Directions in the Theory of Graphs*, pages 239–273. Academic Press, New York, 1973.

M. Rosenblatt. Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472, 1952.

U. Schepsmeier, J. Stöber, and E. C. Brechmann. *Statistical inference of vine copulas*, 2013. Manual for the R package `VineCopula`.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2): 461–464, 1978.

A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231, 1959.

M. Smith, A. Min, C. Almeida, and C. Czado. Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, 105(492):1467–1479, 2010.

K. Song. Testing conditional independence via Rosenblatt transforms. *The Annals of Statistics*, 37(6B):4011–4045, 2009.

P. Spirtes and C. Glymour. An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review*, 9(1):62–72, 1991.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT Press, Cambridge, Massachusetts, second edition, 2000.

J. Stöber, H. Joe, and C. Czado. Simplified pair copula constructions—limits and extensions. arXiv:1205.4844 [stat.ME], 2012.

I. Tsamardinos, L. E. Brown, and C. F. Aliferis. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, 65(1):31–78, 2006.

T. Verma and J. Pearl. Equivalence and synthesis of causal models. In P. Bonissone, M. Henrion, L. Kanal, and J. Lemmer (Eds.), *Uncertainty in Artificial Intelligence*, volume 6, pages 220–227. Elsevier Science, New York, 1991.

Q. H. Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307–333, 1989.

J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. John Wiley & Sons, Chichester, 1990.

S. Wright. Correlation and causation. *Journal of Agricultural Research*, 20(7):557–585, 1921.

K. Zhang, J. Peters, D. Janzing, and B. Schölkopf. Kernel-based conditional independence test and application in causal discovery. In F. Cozman and A. Pfeffer (Eds.), *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, pages 804–813. AUAI Press, Corvallis, Oregon, 2011.

Abbreviations

AIC	Akaike's information criterion
BIC	Bayesian information criterion
C-vine	canonical vine
cdf	cumulative distribution function
cf.	compare (Lat. *confer*)
CG	chain graph
D-vine	drawable vine
DAG	directed acyclic graph
i.e.	that is (Lat. *id est*)
i.i.d.	independent and identically distributed
ML	maximum likelihood
MSE	mean squared error
PCC	pair-copula construction
pdf	probability density function
R-vine	regular vine
TDC	tail dependence coefficient
TSP	travelling salesman problem
UG	undirected graph
w.l.o.g.	without loss of generality